

|CURIEUX|

ACADEMIC REVIEW

February Issue

Issue 4 Part 2

Editing Staff

Chief-In-Editor

George Hasnah

Lead Editor

Amanuel Geremew

Ashley Du

Assisting Editors

Jing Yan Huang

Nidhi Potineni

Sophia Kong

Tony Yuan

Ridit Aggarwal

Manu Sharma

Rishi Gadikota

Eric Wang

Table Of Contents

Page 6: Crisis Communication and Brand Reputation in the Digital Era: A Comparative Study of Corporate Response Strategies in the United States and Kazakhstan By Aidana Orazalina

Page 35: Microglial Activation in Children with Early Cortical Thinning Predicts ADHD Symptoms: Attention, Memory, and Planning Deficits in Adolescence By Vishika Challapalli

Page 50: Fast Radio Bursts: An Introduction to One of Astronomy's Most Mysterious Signals By Rachelle Pinkerton

Page 57: Battling Institutional Racism: The Importance of Activist Messaging in Combatting Social Separation By Mengchen Zhang

Page 67: A Systematic Review of the COVID-19 Lockdown Impacting Children with Autism: Changes in Applied Behavioral Analysis and Quality of Life By Adriana Victoria

Page 79: Economic Impact of the VASEN 2025 Linhai Youxi Ultra Trail Triathlon on Taizhou, Zhejiang Province: A Comparative Analysis with the Tsaigu Trail Race Series By Xinle (Simon) Chen

Page 88: Moral Intuition Should Not Be Trusted By Xinyi Wu

Page 94: A Weapon of Mass Distraction: Misinformation in the Russian-Ukrainian Conflict By Jerome Ong

Page 118: The Link Between Solar UV Radiation and the Development of Skin Cancer Through DNA Damage By Emi Kariqi

Page 135: Descriptive Analysis of NBA Shooting Trends: A Comparative Study of Five Teams in the Most Recent Regular Season By Adish Reddy and Neharika Shrivastava

Page 149: Relationship Between Nighttime Sleep, Daytime Naps, and GPA By Yusuf Abdellatif

Page 155: A Quantitative Model on Physical Activity and Learning Readiness among Preschool Children By Muhun Jung

Page 163: A Comprehensive Review of VEGF and IL-6 Differences in Long COVID Progression Between Sexes By Carissa Chen

Page 175: Symbols of Conflict: How Wartime Propaganda Misappropriated History and Impacted National Identity in the 20th-Century By Ahriyan Anand

Page 192: How Do The Actual Economic Impacts Of the Super Bowl and Coachella Compare To The Implications Set by the Companies and the Media? By Ruchika Varanasi

Page 203: How Do Airdrops Influence Long-Term User Retention in DeFi Ecosystems? By Bassel El Hage

Page 219: CRISPR's Ability: The Different Systems Used to Edit DNA and RNA By Arnav Jagtap

Page 232: Urban Heat Vulnerabilities in Mid-Sized Cities: A Case Study of Lansing, Michigan By Haneesha Padavala

Page 240: Harnessing the cBioPortal Database to Identify Genetic Determinants of Survival in Breast Cancer By Claire Loftus

Page 251: Consumed By Perception: Evaluating How Glorified Diets and Superfoods Reshape Societal Norms By Sydney Pryor

Page 256: An AI-Powered Night Vision System for Pedestrian Detection and Illumination at Crosswalks Using a Custom YOLOv8 Model By Agustya Agustya

Page 273: Nanotechnology In Metastatic Breast Cancer By Darara Meskela

Page 289: Can GPT Generated Poetry, Infused with Extreme Metaphors Be Perceived as More Profound than Original Human Written Poetry? By Devansh Gupta

Page 297: Leveling the Playing Field: A Comparison of Economic Structures and Systems Across Different Sports Leagues By August Berglind

Page 308: Dorsal Fin Edge Proportions and Their Relationship to Swimming Strategy in Sharks By Benjamin Janitz

Page 314: Predicting ecDNA Presence with Machine Learning on Inexpensive Genomic Data By Esha Verma¹, Azhar Khandekar²

Page 324: Review of the Impacts of Carbon Nanotubes on the Freeze-Thaw Resistance of Concrete By Rickie Chen

Page 334: Holding Up a Generation: Reimagining How Schools, Parents, and Peers Support Teen Mental Health By Jiya Aghi, Kaeya Aghi, Sonia Jain

Page 345: Existing Methods for Managing Fear, Anxiety, and Stress in Shelter Animals By Nancy Sun

Page 361: Mathematical Foundations in Hindu Temple Architecture: An Analysis of Geometrical, Proportional and Astronomical Principles in Structural Design and Decorative Motifs By Bhadra Ajayan

Page 373: The Effects of Music Therapy on Cognitive Function in Alzheimer's Disease By Jenna Kim

Page 386: A Nation Ignited: The Origins of Orbit and the Public's Ascent By Sienna Tolani

Page 413: A Cross-Cultural Analysis - Effect That Advertising Has on Socialization Focusing on Skin Tone and the Perception of Women Aanya Nadkarni By Aanya Nadkarni

Page 427: Efficacy of the BNT116 mRNA Vaccine for the Treatment of Non-Small Cell Lung Cancer By Chloe Tam

Page 435: Gender and Higher Education: Exploring the Academic Journeys of Female Students in Bangladesh By Ahsun Abdul Goffar

Page 450: Interdisciplinary Potential of Cellular and Acellular Therapies in Cardiac Regenerative Medicine By James Cooper

Page 456: The Evolution of Neuromorphic Computing: An Analysis of Neuromorphic Systems, Memristive Devices, and Future Challenges By Vivian Chen

Page 464: Fourier-Domain Feature Extraction for Stable Sensory Feedback in Neuro-Haptic Prosthetics By Martin Meneses

Page 476: The Role of Payment for Order Flow in The Rise of Commission-Free Trading Platforms By Julien Cohen

Page 491: The Promise of Fusion Energy and Challenges after Breakeven with Commercial Reactors By Vatshal Gaur

Page 497: The Application of Stem Cells as a Treatment for Glioblastoma By Baek, H.J., Pearson, A.N.

Page 512: The Effect of Minimum Wage Increases on the Local Economy: A Case Study of Seattle By Isaac Sedighi, Rocco Graziosi

Page 521: Advancements in Machine Learning Applications for Earthquake Prediction By Alex Chen

Page 528: Leveraging Existing Nanostore Infrastructure for Inclusive Market Access: Consignment-Based Entrepreneurship in Jordan By Leyth Sharaf

Page 538: The Chemical and Energetic Constraints of Chemosynthesis-Based Ecosystems in Sunless Environments By Neeva Gangopadhyay

Page 542: Substance Abuse and Homelessness, Which Comes First? Is substance abuse the cause of homelessness, or is it the consequence? By Allison Lu

Page 546: The Hidden Cost of Beauty: Shark-Derived Squalene and the Ethics of Cosmetic Transparency By Lilian Thomas

Page 556: Aesculin: Natural Functions, Unique abilities, Current and Future Medical Possibilities By Viacheslav Detkov

Page 577: How Cook County's Complex Bureaucracy Worsens Inequity and Compromises Public Trust By Kayla Moore

Page 589: Utilising pomelo peel (*Citrus grandis*) as a sustainable bio-absorbent to tackle water scarcity: A novel agricultural research By Anwesha Shantanu Choudhari, Rajas Raghav Khare, Sitalakshmi Parameswaran, Revathi Haval, Sheetal Shantanu Choudhari, and Vibhuti Shreesh Mhatre

Page 599: Gut Microbiome By Rishi Jain

Page 602: Housing Scarcity as an Institutional Outcome: Immigration, Capital Flows, and Urban Governance By Matthew Lewis

Page 612: How Does Exposure to War Affect Children's Physical and Mental Health, and What Strategies can be Implemented to Protect Them? By Viktoria H. Gilbert

Crisis Communication and Brand Reputation in the Digital Era: A Comparative Study of Corporate Response Strategies in the United States and Kazakhstan

By Aidana Orazalina

Abstract

Crisis communication is a defining aspect of corporate reputation management in the digital era. As technology has accelerated the pace of information exchange, it has also increased the vulnerability of organizations to reputational harm. This paper explores crisis communication strategies in the United States and Kazakhstan, focusing on how corporations use digital media to maintain public trust and brand legitimacy during crises. Through a comparative analysis of theoretical frameworks, case studies, and cross-cultural practices, this review highlights differences in institutional maturity, transparency norms, and stakeholder engagement. The findings demonstrate that while U.S. corporations emphasize immediacy, transparency, and strategic apology, Kazakh companies are gradually integrating digital communication tools and public dialogue into their crisis management practices. The paper concludes that effective crisis communication in both contexts depends on adaptability, cultural awareness, and ethical engagement in the digital public sphere.

Keywords

Crisis communication, brand reputation, digital media, corporate communication, public relations, transparency, stakeholder trust, Kazakhstan, United States

1. Introduction

Crisis communication has become one of the most studied and strategically essential areas within public relations and corporate communication. In the contemporary digital environment, where information spreads rapidly and reputations can shift within hours, effective crisis management determines whether a company survives reputational damage or collapses under public scrutiny. The concept of a “corporate crisis” encompasses a wide range of incidents, from product recalls and financial fraud to environmental disasters and social media controversies (Coombs, 2015). What unites these diverse scenarios is the communication process that follows, in which organizations attempt to explain, justify, or remediate events that threaten public trust.

In the United States, the practice of crisis communication is supported by decades of research, institutional frameworks, and a mature media environment that expects openness and accountability. Scholars such as W. Timothy Coombs (2014) and Robert L. Heath (2010) have established foundational models like the Situational Crisis Communication Theory (SCCT), which guide organizations in crafting context-sensitive responses. Companies like Johnson & Johnson, Boeing, and Starbucks have become case studies in the success or failure of crisis response strategies. The digital transformation of media has added new complexity, requiring

corporations to monitor online sentiment and engage in real-time responses to viral narratives (Jin, Liu, & Austin, 2014).

In Kazakhstan and the broader Central Asian context, the field of crisis communication is still developing. As Kazakhstan's economy diversifies and integrates into global markets, corporations are encountering new expectations for transparency, stakeholder dialogue, and corporate social responsibility (Tulegenova & Sadykova, 2022). Traditional top-down communication models, common in post-Soviet contexts, are gradually giving way to more interactive, audience-centered approaches driven by digital media. However, institutional and cultural factors continue to influence how crises are managed and how public trust is restored. This paper aims to examine how corporations in the United States and Kazakhstan approach crisis communication and brand reputation management in the digital age. By analyzing academic research, industry reports, and notable case studies, this review identifies key similarities and differences in communication styles, strategic responses, and public engagement mechanisms. The study also seeks to highlight the role of cultural, technological, and regulatory contexts in shaping crisis management outcomes.

The review is organized into several sections. First, it presents a detailed literature review outlining major theories and models of crisis communication. Next, it discusses the methodological scope and parameters of the comparative analysis. Subsequent sections examine crisis communication practices in the United States and Kazakhstan, followed by a comparative evaluation of both contexts. Finally, the paper discusses emerging challenges in digital reputation management and proposes directions for future research.

2. Literature Review

The field of crisis communication has evolved over several decades, transitioning from reactive damage control to a proactive and strategic function within corporate communication. The academic foundation of crisis communication rests on the idea that organizations are not only responsible for responding to crises but also for shaping narratives and perceptions that influence stakeholder trust. This section reviews the key theoretical perspectives, models, and contemporary discussions in the field, with special emphasis on the role of digital communication and cultural context.

2.1 Foundational Theories in Crisis Communication

One of the most influential theoretical models in the study of crisis communication is the Situational Crisis Communication Theory (SCCT), developed by W. Timothy Coombs (2007, 2015). The SCCT framework posits that the effectiveness of a crisis response depends on the type of crisis and the degree of organizational responsibility perceived by the public. According to Coombs (2014), crisis managers must align their communication strategy with the level of reputational threat, adopting approaches such as denial, justification, apology, or corrective action based on situational needs. This theoretical foundation has guided much of contemporary research and is considered the cornerstone of evidence-based crisis management.

Similarly, the Image Repair Theory developed by William L. Benoit (1995) has been instrumental in understanding how organizations attempt to restore reputation following a crisis. Benoit's typology includes five broad strategies: denial, evasion of responsibility, reduction of offensiveness, corrective action, and mortification. These strategies emphasize the rhetorical dimensions of corporate apology and responsibility. The theory has been widely applied to corporate case studies such as Exxon's handling of the Valdez oil spill and Toyota's product recall crises.

Another key perspective is the Restorative Rhetoric Model proposed by Keith Michael Hearit (2006), which extends beyond immediate crisis response to long-term reputation recovery. This model emphasizes transparency, accountability, and ethical consistency in communication. Hearit (2006) argues that authentic corporate communication following a crisis should focus on moral reconstruction rather than mere image restoration.

Together, these frameworks underline a shared principle: crisis communication is both a strategic and ethical act. The goal is not simply to manage damage but to rebuild legitimacy and trust, which are fundamental to corporate survival in the digital era.

2.2 The Digital Transformation of Crisis Communication

The advent of digital and social media has fundamentally altered the landscape of corporate communication. Scholars agree that social media platforms, particularly Twitter, Facebook, Instagram, and TikTok, have redefined the speed and visibility of crisis events (Jin et al., 2014; Liu, Jin, & Austin, 2018). Unlike traditional media, where information flow was linear and controlled, digital media have created a decentralized and participatory communication environment.

According to Austin and Jin (2017), the Social-Mediated Crisis Communication (SMCC) Model describes how information during crises flows among organizations, the public, and influencers. This model identifies three actors: information creators (organizations and publics), information carriers (media and influencers), and information consumers (audiences). The dynamic interplay among these groups determines the credibility, reach, and impact of crisis narratives.

In the U.S. context, the digital shift has encouraged corporations to adopt real-time monitoring and data analytics for crisis detection and response (Fearn-Banks, 2016). Corporate communication teams use sentiment analysis, social listening tools, and influencer engagement to shape responses that align with public expectations. Transparency and immediacy are now considered central to successful crisis communication (Jiang & Luo, 2018). However, the digital environment also presents challenges. The virality of misinformation, the fragmentation of audiences, and the rise of cancel culture have increased the risks associated with corporate crises. Coombs and Holladay (2018) argue that the "social amplification of risk" through digital media can intensify crises beyond their actual scope, making early and consistent communication essential.

In Kazakhstan, digital transformation in communication has followed a different trajectory. The use of social media platforms such as Instagram, Telegram, and TikTok has expanded rapidly since 2015, especially among younger audiences. According to Kantar Central Asia (2022), nearly 70 percent of Kazakh Internet users engage with brands through social media. However, corporate crisis communication practices remain less institutionalized compared to the U.S. Public relations departments in Kazakh companies are still evolving, often focusing on reputation monitoring rather than proactive engagement (Tulegenova & Sadykova, 2022). This suggests that digitalization is outpacing the development of structured crisis management frameworks in emerging economies.

2.3 Corporate Reputation and Stakeholder Trust

Corporate reputation is defined as the collective perception of a company's integrity, reliability, and competence over time (Fombrun & van Riel, 2004). Reputation is not a static asset but a dynamic construct shaped by communication and experience. During crises, trust becomes both the most fragile and the most vital element of stakeholder relationships (Dowling, 2016).

According to Kim, Avery, and Lariscy (2009), reputation repair depends heavily on the perceived sincerity and consistency of corporate responses. In the digital era, where stakeholders can verify information instantly, authenticity has become a critical success factor. The Trust Recovery Framework developed by Gillespie and Dietz (2009) emphasizes the need for both cognitive trust (belief in competence) and affective trust (belief in goodwill). Corporations that communicate empathy and responsibility are more likely to restore both dimensions of trust. In comparative studies, researchers have found that cultural context influences how trust is built and maintained (Huang, 2000). In the United States, transparency, accountability, and direct apology are often expected as signs of responsibility. In contrast, in high-context cultures such as Kazakhstan, indirect communication and maintaining social harmony may take precedence over public admission of fault. This cultural dimension affects how organizations frame their crisis messages and how stakeholders interpret them (Hofstede, 2011).

2.4 The Role of Corporate Ethics and Social Responsibility

Ethical communication has emerged as a central concern in crisis management. Coombs and Holladay (2012) argue that ethical communication goes beyond compliance with corporate policy and includes honesty, accountability, and respect for affected stakeholders. Ethical lapses in crisis response, such as withholding information or blaming others, can lead to long-term reputational damage.

Corporate Social Responsibility (CSR) is closely linked to crisis communication because it frames public expectations of a company's moral stance. Companies with strong CSR reputations often have a "reputation reservoir" that cushions the impact of crises (Kim & Kim, 2019). For example, in the U.S., brands like Patagonia and Ben & Jerry's are often forgiven faster during controversies due to their consistent ethical branding. In Kazakhstan, CSR practices

are growing in importance, particularly in industries such as energy, banking, and telecommunications, where public scrutiny is increasing (Mukhamedova & Beknazarov, 2021).

2.5 Cross-Cultural Communication and Crisis Response

Culture is a decisive variable in understanding how crisis communication strategies are perceived. According to Hofstede's (2011) cultural dimensions, the U.S. exhibits low power distance and high individualism, encouraging direct communication and accountability. Kazakhstan, by contrast, scores higher on power distance and collectivism, which translates into more hierarchical decision-making and cautious public messaging.

Research by Kim and Niederdeppe (2013) suggests that in collectivist societies, preserving group harmony and avoiding public confrontation are central to communication strategies. Consequently, Kazakh companies may prioritize internal resolution and delayed public statements, while U.S. firms often engage in early, public-facing responses. This difference reflects the broader societal expectations regarding authority, responsibility, and openness.

Cross-cultural PR research highlights that Western models such as SCCT need adaptation to fit local contexts (Taylor, 2010). Localized crisis management models must consider cultural values, media systems, and institutional trust levels. In Kazakhstan, where public confidence in corporate and governmental institutions is still developing, crisis communication must balance transparency with cultural sensitivity.

3. Theoretical Framework and Methodology

3.1 Theoretical Foundation

This research review is guided by three principal theoretical frameworks that together provide a comprehensive understanding of corporate crisis communication in a global context: Situational Crisis Communication Theory (SCCT), the Social-Mediated Crisis Communication (SMCC) Model, and Cross-Cultural Communication Theory. Each of these perspectives contributes to explaining how corporations craft, deliver, and adapt messages to diverse audiences during crises.

Situational Crisis Communication Theory (SCCT)

Developed by W. Timothy Coombs (2007), SCCT is one of the most widely applied models in the field of crisis communication. It is based on attribution theory, which posits that stakeholders evaluate organizational responsibility based on perceived causes and responses. According to Coombs (2014), crisis response strategies must be aligned with the degree of responsibility attributed to the organization. SCCT categorizes crises into three primary clusters: victim crises (where the organization is also a victim, such as in natural disasters or rumors), accidental crises (where the event was unintentional but linked to organizational actions), and preventable crises (where negligence or wrongdoing is evident).

The theory emphasizes matching response strategies to the reputational threat. For example, in victim crises, organizations may use strategies of denial or scapegoating, while in preventable crises, strategies involving apology and corrective action are more effective (Coombs & Holladay, 2012). The SCCT model also underscores the importance of prior reputation and crisis history, which can either buffer or amplify public reactions. This model has been applied extensively in U.S. corporate case studies, such as BP's oil spill and Toyota's recall crisis, and it provides a structured foundation for comparative analysis with emerging markets like Kazakhstan.

Social-Mediated Crisis Communication (SMCC) Model

The SMCC Model developed by Jin, Liu, and Austin (2014) expands SCCT by integrating the digital media environment into crisis communication theory. It explains how information flows through social media ecosystems, identifying three primary actors: information creators (such as organizations, media, and influencers), information carriers (those who redistribute messages), and information consumers (the public). The model posits that trust and credibility are distributed across these actors, creating complex patterns of influence.

In practice, the SMCC model suggests that organizations must manage both direct communication (official statements) and indirect communication (public interpretation and influencer narratives). According to Austin and Jin (2017), audience engagement during a crisis depends not only on message clarity but also on the organization's responsiveness and emotional tone. This framework is particularly relevant for the U.S. context, where corporations must navigate high-speed public discourse and where misinformation can spread rapidly.

For Kazakhstan, the SMCC model highlights challenges related to media literacy, regulatory frameworks, and the still-evolving relationship between corporations and online publics. The model's flexibility allows for adaptation to contexts where institutional trust and digital infrastructure are in different stages of development.

Cross-Cultural Communication Framework

Culture fundamentally shapes how crises are perceived and how organizations communicate in response. The cross-cultural perspective used in this paper draws primarily on Hofstede's cultural dimensions (Hofstede, 2011) and Hall's high-context and low-context communication model (Hall, 1976). These frameworks help explain why identical communication strategies may yield different outcomes across cultures.

In Hofstede's model, the United States scores high on individualism and low on power distance, reflecting a culture that values transparency, accountability, and direct communication. Kazakhstan, by contrast, exhibits higher power distance and collectivism, indicating hierarchical structures and a greater emphasis on maintaining harmony (Makhmutova & Sadykova, 2020). These dimensions suggest that crisis messages in Kazakhstan may rely more on formal statements and controlled messaging, while U.S. companies are expected to provide immediate and open responses.

By integrating these frameworks, this paper positions crisis communication not as a uniform global practice but as a culturally adaptive process. The combination of SCCT, SMCC, and cross-cultural theories allows for a nuanced comparative analysis that considers both structural and sociocultural factors.

3.2 Methodological Approach

As a review article, this study adopts a qualitative comparative research design. The objective is not to test hypotheses but to synthesize existing academic and industry research on crisis communication strategies across two national contexts. The methodology is divided into three stages: literature selection, case study identification, and comparative thematic analysis.

Literature Selection

The literature for this review was collected from peer-reviewed journals, including *Public Relations Review*, *Journal of Public Relations Research*, *Corporate Communications: An International Journal*, and *Journal of Business Communication*. Industry reports from Edelman (2023), Weber Shandwick (2022), and regional communication agencies in Kazakhstan were also examined to provide real-world insights. The selection focused on publications from 2010 to 2024, ensuring relevance to the digital communication era.

Search keywords included “crisis communication,” “digital media,” “corporate reputation,” “cross-cultural PR,” “brand trust,” and “Kazakhstan corporate communication.” A total of 63 sources were initially reviewed, with 38 selected for inclusion based on methodological rigor and relevance to comparative analysis.

Case Study Identification

To provide contextual grounding, this study reviews several major corporate crises from both the United States and Kazakhstan. For the U.S., cases include Boeing’s 737 MAX crisis (2018–2020), Nike’s social activism campaigns (2018), and Johnson & Johnson’s Tylenol recall (1982) as historical reference points. For Kazakhstan, relevant cases include Air Astana’s safety incident communications (2021), Kaspi.kz’s data security responses (2022), and KazMunayGas’s environmental accountability statements (2023).

These cases were selected to reflect diversity in industry, communication approach, and crisis type. They serve not as empirical samples but as illustrations of how cultural, technological, and institutional differences shape communication strategies.

Comparative Thematic Analysis

The review applies a comparative thematic approach, identifying patterns and contrasts across four analytical dimensions:

1. Crisis response timing and transparency
2. Message framing and tone
3. Stakeholder engagement and digital dialogue
4. Ethical accountability and long-term reputation repair

Each dimension is examined in relation to both the U.S. and Kazakh contexts, guided by theoretical models. The analysis seeks to identify not only best practices but also cultural and systemic limitations that affect communication effectiveness.

This approach aligns with comparative public relations scholarship, which encourages examining non-Western practices as distinct yet evolving frameworks (Sriramesh & Vercic, 2019). By focusing on Kazakhstan, an emerging economy with rapid digitalization, this study contributes to globalizing the understanding of crisis communication beyond Western-centric paradigms.

3.3 Research Validity and Limitations

As a review-based study, this research relies on secondary data and interpretive analysis rather than primary data collection. The validity of findings depends on the credibility of selected sources and the representativeness of analyzed cases. The inclusion of both peer-reviewed academic studies and verified industry reports ensures balance between theoretical rigor and practical relevance.

However, several limitations are acknowledged. First, crisis communication research in Kazakhstan is still developing, which limits the availability of localized academic sources. Second, cultural comparisons inherently risk overgeneralization; thus, this study emphasizes contextual interpretation rather than deterministic conclusions. Finally, the dynamic nature of digital communication means that crisis strategies evolve rapidly, making continual updates necessary for ongoing relevance.

3.4 Ethical Considerations

The review adheres to academic integrity principles, ensuring accurate citation of all sources and fair representation of diverse scholarly perspectives. Ethical analysis also extends to the discussion of corporate behavior, recognizing that crisis communication is not merely a strategic tool but a form of ethical dialogue with the public. Following Bowen (2008), ethical crisis communication requires accountability, empathy, and transparency, which are crucial to rebuilding stakeholder trust.

4. Crisis Communication Strategies in the United States

The United States has long served as the intellectual and practical center of modern crisis communication. Corporate responses to crises in this context are shaped by democratic media traditions, legal accountability, and stakeholder expectations for transparency. Scholars such as Coombs (2015), Fearn-Banks (2016), and Heath (2010) note that the American model of crisis management is built on the premise that swift, truthful, and empathetic communication can mitigate reputational damage. This section reviews how U.S. corporations have applied these principles in practice, focusing on three representative cases: Boeing's 737 MAX crisis, Nike's brand activism campaigns, and Johnson & Johnson's Tylenol recall. Each case illustrates evolving patterns of digital strategy, ethical framing, and public trust recovery.

4.1 Boeing and the 737 MAX Crisis

The Boeing 737 MAX crisis remains one of the most studied corporate failures in recent years. Following two fatal crashes between 2018 and 2019, Boeing faced intense global scrutiny over safety protocols, transparency, and ethical responsibility. According to Elliott and Hanna (2020), Boeing's initial crisis response was marked by hesitation, fragmented messaging, and a perceived lack of empathy toward victims and the flying public. The company's early statements focused on technical explanations rather than human consequences, which intensified negative sentiment across media platforms.

From a theoretical perspective, Boeing's response violated several principles of the Situational Crisis Communication Theory (SCCT). The crisis was clearly preventable, involving design flaws and insufficient pilot training, which meant stakeholders expected a full admission of responsibility and corrective action. Instead, Boeing relied heavily on defensive messaging that aimed to protect its corporate image (Coombs & Holladay, 2012). Public trust eroded further when internal communications revealed that executives had minimized safety concerns.

Digital media amplified these failures. Social media monitoring by Pew Research (2019) showed a sustained wave of online criticism lasting over six months, fueled by influencers and journalists. Boeing eventually shifted its strategy by issuing a formal apology, grounding the fleet, and committing to transparency in its software redesign. However, as Shin et al. (2021) observed, the delay in assuming responsibility resulted in long-term reputational decline and market loss. The case demonstrates that in the U.S. context, transparency and empathy are indispensable, and delayed communication can transform operational crises into existential ones.

4.2 Nike and Brand Activism: The Colin Kaepernick Campaign

Nike's 2018 "Dream Crazy" campaign featuring Colin Kaepernick marked a turning point in corporate communication strategy, blurring the line between crisis management and proactive brand activism. Kaepernick, a former NFL player, had become a controversial figure after kneeling during the national anthem to protest racial injustice. Nike's decision to feature him in a major advertisement sparked immediate backlash, with calls for boycotts across conservative segments of the U.S. population. However, it also earned admiration from younger, socially conscious consumers.

Unlike reactive crisis responses, Nike's campaign exemplified preemptive crisis framing, where the company anticipated controversy and strategically integrated it into its brand narrative (Parganas & Anagnostopoulos, 2020). Through digital storytelling, Nike reinforced its core values of diversity, courage, and social responsibility. The company's communication team employed real-time engagement across Twitter, Instagram, and YouTube, emphasizing empowerment and inclusion rather than apology.

Data analytics by Edelman (2019) later confirmed that Nike's brand favorability among its target demographic increased by 10 percent within a month of the campaign launch. The company's stock price rose, and the campaign won multiple global advertising awards. From an SCCT standpoint, Nike effectively reframed potential reputational risk as moral leadership. As

Carr and Hayes (2021) argue, the campaign illustrates a new paradigm in which companies leverage social values as a resilience mechanism, demonstrating that authenticity can convert controversy into loyalty when communication aligns with brand identity.

4.3 Johnson & Johnson and the Tylenol Recall: The Foundational Model

Although nearly four decades old, the Johnson & Johnson Tylenol case from 1982 remains a seminal example of effective crisis communication. After seven consumers died from cyanide-laced Tylenol capsules, the company faced a life-threatening brand crisis. Instead of minimizing the issue, Johnson & Johnson immediately recalled over 31 million bottles, suspended advertising, and fully cooperated with law enforcement (Benoit, 1995). The company's chairman at the time, James Burke, personally addressed the public through press conferences and television interviews, conveying empathy and moral responsibility.

From a theoretical standpoint, Johnson & Johnson's actions exemplified the corrective action and mortification strategies of Benoit's Image Repair Theory (1995). The company assumed full responsibility, even though it was technically a victim of tampering. According to Fearn-Banks (2016), this approach established a lasting benchmark for ethical corporate behavior, often cited in public relations education as the gold standard of crisis response.

The long-term outcome validated this strategy: Tylenol regained its market share within a year, and Johnson & Johnson's reputation for integrity was strengthened. In a comparative context, the Tylenol case underscores the enduring effectiveness of transparent and ethical communication. As Coombs (2015) notes, the case demonstrates that stakeholder trust is not merely restored through words but through consistent, value-driven action.

4.4 The Role of Digital Media and Real-Time Engagement

Modern U.S. corporations now operate in a media environment where crises are often initiated, amplified, and resolved through digital channels. According to Liu et al. (2018), the rise of social media has democratized public relations, shifting control of narratives from corporations to audiences. Platforms such as Twitter and Reddit have become critical arenas for crisis discourse, where misinformation can spread rapidly and influence investor confidence. Companies increasingly rely on digital listening tools and data-driven response models to manage reputational risk (Wigley & Zhang, 2017). Real-time monitoring allows for faster identification of emerging crises, while sentiment analysis informs tone and message adaptation. The 2020 Edelman Trust Barometer revealed that 81 percent of U.S. consumers expect companies to respond to crises within 24 hours of their occurrence.

However, immediacy also carries risks. Hastily issued statements can appear insincere or inconsistent if not grounded in verified facts (Jiang & Luo, 2018). The challenge lies in balancing speed with accuracy, a dilemma often referred to as the "trust-velocity paradox." Successful corporations integrate cross-functional crisis teams that combine legal, communications, and digital analytics expertise to maintain coherence and credibility.

4.5 Ethical Accountability and Transparency

Ethics remains central to effective crisis management. In the United States, public trust is strongly tied to perceptions of honesty and moral responsibility. Bowen (2008) emphasizes that ethical communication is not only a moral obligation but also a strategic advantage, as it reinforces stakeholder loyalty. The U.S. media landscape, characterized by investigative journalism and active citizen oversight, imposes additional pressure on corporations to maintain transparency.

Boeing's delayed acknowledgment of fault, for instance, contrasted sharply with Johnson & Johnson's immediate transparency. Similarly, companies like Starbucks and United Airlines have faced public backlash when perceived as prioritizing image over accountability. As noted by Heath (2010), sustainable crisis communication requires consistency between declared values and observed actions.

U.S. corporations have thus institutionalized ethical guidelines within their crisis management plans, often codified in corporate governance frameworks. For example, the Public Relations Society of America (PRSA) Code of Ethics outlines principles such as advocacy, honesty, and fairness as non-negotiable standards (PRSA, 2022). These ethical codes serve not only as professional benchmarks but as mechanisms of public reassurance during crises.

5. Crisis Communication Strategies in Kazakhstan

Crisis communication in Kazakhstan represents a dynamic and developing field that reflects the intersection of traditional communication norms, evolving digital ecosystems, and globalization of corporate practices. While the United States has a well-established theoretical and institutional base for crisis management, Kazakhstan's approach remains in a formative stage. The country's communication strategies are deeply influenced by its post-Soviet administrative culture, hierarchical corporate governance, and the gradual integration of international public relations (PR) frameworks (Tulegenova & Sadykova, 2022).

In the past decade, rapid digitalization, social media proliferation, and a growing middle class have created new expectations for transparency and responsiveness. According to Kantar Central Asia (2022), more than 75 percent of Kazakh citizens actively use social media platforms such as Instagram, Telegram, and TikTok. This digital transformation has reshaped public discourse, compelling companies to engage with audiences in real time. However, despite these advancements, corporate crisis communication often remains reactive rather than preventive, and institutional trust in both public and private sectors continues to evolve.

This section explores key aspects of Kazakhstan's crisis communication landscape through three representative cases: Air Astana's safety incident communication, Kaspi.kz's data security management, and KazMunayGas's environmental accountability response. These cases highlight the growing influence of digital communication, emerging transparency norms, and the adaptation of global crisis management models to local conditions.

5.1 Air Astana and Safety Incident Communication

As Kazakhstan's flagship airline, Air Astana serves as one of the most prominent examples of professional corporate communication in Central Asia. In 2021, an incident involving a technical failure during a domestic flight drew public concern and significant media attention. Although no casualties occurred, the situation required an immediate communication response to reassure both passengers and regulators.

Air Astana's crisis response followed a relatively structured process. Within hours of the incident, the airline issued an official statement on its website and social media platforms, providing factual details and confirming the safety of all passengers. This approach reflected elements of Situational Crisis Communication Theory (SCCT), particularly the emphasis on transparency and corrective action (Coombs, 2014). The airline also announced a technical investigation and temporarily suspended the aircraft model involved in the event.

However, qualitative analysis of public comments on Telegram and local news portals revealed skepticism regarding the company's initial tone, which some audiences perceived as overly formal and detached (Tulegenova & Sadykova, 2022). This reaction illustrates the cultural challenge of balancing professionalism with empathy in Kazakhstan's communication environment. While Air Astana succeeded in managing factual accuracy, it faced difficulty establishing emotional connection with the public. The company later adjusted its strategy by sharing video messages from senior executives, providing reassurance and emphasizing a culture of safety.

This case suggests that crisis communication in Kazakhstan is gradually evolving from information control toward stakeholder dialogue. Air Astana's approach aligns with international best practices in content structure but demonstrates the continuing need for cultural adaptation to increase authenticity and emotional resonance.

5.2 Kaspi.kz and Data Security Management

Kaspi.kz, one of Kazakhstan's largest fintech companies, has become a benchmark for digital innovation and brand trust in Central Asia. Its crisis communication capacity was tested in 2022 when rumors about a potential data breach circulated on social media, prompting public anxiety regarding account security. The situation did not involve an actual security failure, yet it represented a reputational crisis in the digital sphere.

Kaspi.kz responded within hours, issuing an official statement on Telegram and Instagram that denied the rumors, reaffirmed the stability of its systems, and provided contact details for direct customer verification. According to analysts at PwC Kazakhstan (2022), the company's quick, transparent, and fact-based messaging effectively limited misinformation spread. Kaspi.kz also used influencers and digital ambassadors to disseminate clarifications, following principles of the Social-Mediated Crisis Communication (SMCC) Model (Jin, Liu, & Austin, 2014).

The company's digital responsiveness mirrors Western crisis communication practices and demonstrates the growing sophistication of corporate communication in Kazakhstan's private sector. However, unlike many U.S. corporations, Kaspi.kz did not issue an apology,

which reflects differences in cultural expectations regarding accountability. In high-context communication environments like Kazakhstan, direct admissions of responsibility may be interpreted as weakness or guilt rather than transparency (Hofstede, 2011). Consequently, companies often prefer controlled messaging that emphasizes reassurance rather than apology. Kaspi.kz's case demonstrates that digital communication infrastructure in Kazakhstan has matured significantly, but communicative culture still retains elements of formality and cautiousness. The company's ability to control narrative through its own channels, rather than relying solely on traditional media, marks a notable step toward modern crisis communication practices.

5.3 KazMunayGas and Environmental Accountability

KazMunayGas (KMG), Kazakhstan's state-owned oil and gas company, offers a contrasting example of crisis communication in a heavily regulated and politically sensitive sector. In 2023, an environmental incident involving an oil leak in the Caspian region drew criticism from environmental groups and international observers. The company's communication response provides insight into how public-sector corporations balance transparency with political and institutional constraints.

KMG initially released a formal statement emphasizing containment efforts and technical measures, citing collaboration with the Ministry of Ecology. The communication adopted an informational rather than dialogic tone, focusing on operational recovery rather than public engagement. Analysts at the Kazakhstan Institute for Strategic Studies (2023) noted that while the company provided regular updates, its messaging lacked emotional empathy and acknowledgment of ecological impact.

From a theoretical perspective, KMG's approach aligns with the instructing information and adjusting information stages of Coombs' (2015) SCCT but falls short in the area of reputation repair. The company prioritized institutional credibility and compliance over relational trust-building. Moreover, cultural and bureaucratic factors limited the speed of message dissemination, resulting in perceptions of delayed communication (Mukhamedova & Beknazarov, 2021).

Comparatively, KMG's communication model differs significantly from Western practices that emphasize proactive public apology and visible CEO accountability. Nonetheless, recent reforms in Kazakhstan's corporate governance and environmental reporting indicate gradual movement toward global standards. The KMG case underscores the complexity of managing crisis communication in state-linked enterprises, where political context and public perception intersect.

5.4 Patterns in Kazakh Corporate Communication

Across these cases, several patterns emerge that reflect the transitional nature of Kazakhstan's corporate communication environment:

1. Most companies tend to respond after a crisis unfolds rather than implementing proactive risk communication. This reflects both organizational culture and limited institutional experience in strategic PR management.
2. Corporate communication often prioritizes factual accuracy and official tone, with limited emphasis on emotional connection. While this reinforces professionalism, it can reduce perceived empathy, especially in consumer-facing sectors.
3. The use of social media platforms for official communication is expanding rapidly. Companies increasingly rely on Telegram, Instagram, and Facebook for public statements, bypassing traditional press channels.
4. Indirect communication styles, preference for hierarchical approval, and avoidance of public confrontation influence how crises are framed and addressed. These cultural factors align with high-context communication norms described by Hall (1976) and Hofstede (2011).
5. Both private and state-owned corporations are adopting elements of Western crisis management frameworks, often through international partnerships or consultant-led training. This trend indicates a growing recognition of PR as a strategic management function rather than a reactive publicity tool.

5.5 Emerging Trends and Challenges

The future of crisis communication in Kazakhstan depends largely on three interconnected developments: digital transformation, regulatory modernization, and cultural adaptation.

First, digital transformation continues to redefine how information circulates in society. The increasing influence of online journalism and citizen reporting creates both opportunities and risks. Companies that actively monitor social media and engage with audiences in real time are better positioned to mitigate misinformation (Kantar Central Asia, 2022).

Second, regulatory modernization is reshaping the relationship between corporations and the public. Recent transparency requirements under Kazakhstan's Environmental Code (2021) and financial reporting reforms are pushing companies toward greater disclosure. These developments are expected to enhance the institutional framework for public accountability.

Third, cultural adaptation remains a gradual process. While Western PR models emphasize direct apology and open dialogue, Kazakh corporations often navigate complex social hierarchies where public humility may be misunderstood. As public expectations evolve, companies will need to balance cultural sensitivity with international norms of authenticity and transparency.

In conclusion, Kazakhstan's crisis communication landscape illustrates both progress and persistent gaps. The adaptation of global frameworks such as SCCT and SMCC demonstrates growing strategic awareness, yet effective implementation requires deeper cultural contextualization and stronger institutional support.

6. Comparative Analysis: Cultural and Institutional Differences

Crisis communication practices in the United States and Kazakhstan reveal both convergence and divergence that stem from structural, cultural, and technological differences. The comparison highlights how global frameworks such as Situational Crisis Communication Theory (SCCT) and the Social-Mediated Crisis Communication (SMCC) Model are interpreted and operationalized within distinct sociocultural environments. Although globalization has encouraged the diffusion of best practices in public relations, crisis communication remains profoundly shaped by national culture, media systems, and public expectations.

6.1 Institutional Maturity and Organizational Readiness

The United States demonstrates a high degree of institutional maturity in both corporate communication and public relations infrastructure. According to Coombs (2015), crisis communication in the U.S. is typically integrated into corporate governance and risk management systems, supported by specialized communication departments and standardized response protocols. Many corporations conduct regular crisis simulations, employ professional public relations consultants, and use data-driven monitoring systems to anticipate potential risks (Heath, 2010).

In contrast, Kazakhstan's corporate communication systems are relatively young. Many organizations only began formalizing public relations departments within the last two decades, and crisis communication as a strategic discipline is still developing (Tulegenova & Sadykova, 2022). Although companies such as Kaspi.kz and Air Astana demonstrate growing professionalism, the overall institutional framework lacks standardized training, accreditation systems, and regulatory oversight comparable to that in the U.S. This institutional gap often leads to inconsistent crisis management practices and slower response times.

Institutional readiness directly influences communication outcomes. In the U.S., companies are prepared to respond quickly and confidently because protocols are embedded into their organizational culture. In Kazakhstan, hierarchical decision-making and bureaucratic approval processes can delay public statements, reducing the effectiveness of real-time engagement (Mukhamedova & Beknazarov, 2021). This difference underscores the importance of developing localized training programs that combine global expertise with national context.

6.2 Cultural Dimensions and Communication Style

Cultural context remains one of the most significant variables explaining differences in crisis response. Drawing on Hofstede's (2011) cultural dimensions, the United States exhibits high individualism, low power distance, and a preference for direct communication. These characteristics promote transparency, public accountability, and openness in addressing crises. American companies are expected to respond promptly, admit mistakes, and articulate corrective measures clearly.

Kazakhstan, in contrast, demonstrates higher power distance and collectivism, which foster respect for authority, hierarchical communication, and avoidance of public confrontation

(Hall, 1976). In this environment, organizations may prioritize internal problem-solving before addressing the public, valuing harmony and reputation protection over immediate disclosure. This tendency explains why companies such as KazMunayGas and Air Astana often rely on formal, carefully worded statements rather than spontaneous digital interaction.

These cultural dynamics also influence emotional tone. U.S. corporations frequently employ empathetic and conversational language, often featuring CEOs or spokespersons directly addressing audiences through video or social media. In Kazakhstan, emotional appeals are less common and may be viewed as unprofessional. The preference for formal tone reflects broader societal values related to authority and decorum. As Kazakh society becomes more digitally engaged, this cultural paradigm may gradually shift toward greater openness and personalization.

6.3 Media Systems and Public Expectations

Media structure and public access to information strongly shape crisis communication outcomes. The United States has a long tradition of independent journalism and investigative reporting, which creates strong external pressure on corporations to disclose information quickly and accurately (Fearn-Banks, 2016). The presence of watchdog media and a legally protected free press fosters transparency as a normative expectation.

In Kazakhstan, media systems operate under a more hybrid model, combining state influence with private and digital media growth (Kantar Central Asia, 2022). Traditional outlets often focus on official narratives, while social media platforms such as Telegram and Instagram have become primary sources of real-time information for the public. This shift creates both opportunities and risks. On one hand, social media enables companies to bypass traditional gatekeepers and engage directly with audiences. On the other hand, the rapid spread of rumors and limited institutional experience with online crisis management can lead to miscommunication.

Public expectations also differ across contexts. U.S. stakeholders expect transparency and timely engagement, often demanding corporate accountability within hours of a crisis. In Kazakhstan, audiences are more accustomed to formal, staged communication. While this expectation is gradually changing among younger, digitally literate citizens, the generational divide in media trust remains significant. Consequently, corporations must balance the expectations of traditional stakeholders with the demands of an increasingly digital public.

6.4 Ethical and Reputational Considerations

Ethical accountability forms the foundation of crisis communication. In the U.S., ethics is both a moral and strategic requirement. Bowen (2008) argues that ethical transparency enhances stakeholder loyalty and strengthens long-term brand equity. Companies such as Johnson & Johnson and Starbucks have shown that proactive ethical responsibility can transform crises into opportunities for renewed trust. Ethical communication in the U.S. context typically involves immediate disclosure, direct apology, and visible corrective action.

In Kazakhstan, ethical expectations are evolving. Public relations as a professional field is still developing standards related to ethics and corporate responsibility (Tulegenova & Sadykova, 2022). Many companies still equate ethical communication with compliance or information control rather than public dialogue. Cultural factors, such as deference to authority and risk aversion, further limit the adoption of open ethical disclosure. However, recent policy reforms and increasing collaboration with international organizations are promoting greater awareness of corporate ethics.

This divergence has implications for reputation management. U.S. corporations often treat reputation as an intangible asset measured through indices and stakeholder surveys (Fombrun & van Riel, 2004). In Kazakhstan, reputation is often associated with public image and state alignment rather than multidimensional trust metrics. As globalization progresses, Kazakh corporations are beginning to adopt Western models of reputation assessment, which emphasize transparency, stakeholder dialogue, and sustainability communication.

6.5 Adaptation of Global Frameworks

Both contexts demonstrate the applicability of SCCT and SMCC frameworks but with different levels of institutional and cultural adaptation. In the U.S., these theories are applied almost seamlessly, with corporate communication teams trained in their use. SCCT principles guide the alignment of message type with crisis responsibility, while the SMCC model informs digital monitoring and audience engagement strategies (Coombs, 2015; Jin, Liu, & Austin, 2014).

In Kazakhstan, these frameworks are selectively applied. Corporations often adopt the structural elements of SCCT, such as issuing official statements and corrective actions, but adapt emotional and relational elements to fit local norms. The SMCC model is less institutionalized due to limited social media analytics infrastructure and fewer professional training opportunities. Nevertheless, companies such as Kaspi.kz illustrate successful adaptation by employing real-time engagement and influencer-based messaging to counter misinformation. Cross-cultural adaptation is therefore not about replication but contextual translation. Effective crisis communication requires not only technical adherence to global models but also sensitivity to local values and expectations. This balance defines the next stage of professionalization in Kazakhstan's corporate communication landscape.

6.6 Toward Convergence and Global Integration

Despite evident differences, both the United States and Kazakhstan are moving toward convergence in crisis communication practices due to globalization and digital interconnectivity. Multinational corporations operating in Kazakhstan, such as Chevron and Coca-Cola, have introduced global standards of transparency and stakeholder engagement, influencing local corporate culture (Mukhamedova & Beknazarov, 2021). At the same time, U.S. companies increasingly recognize the need to adapt messages to diverse cultural audiences, emphasizing global-local communication balance.

The integration of technology and social listening tools is fostering a universal trend toward data-informed crisis management. Both contexts show a shift from reactive communication to predictive and preventive strategies. Digital transformation is bridging institutional gaps by making information more accessible and empowering publics to demand accountability.

However, complete convergence remains unlikely in the near term. Cultural norms, regulatory environments, and political contexts continue to shape communication priorities. The U.S. model will likely remain centered on transparency and individual responsibility, while Kazakhstan's approach will continue to balance hierarchical structure with gradual democratization of information flow.

7. Digital Media, Trust, and Brand Reputation

Digital media have transformed crisis communication into a dynamic, continuous, and participatory process. Unlike in traditional public relations, where corporations controlled the flow of information, modern communication ecosystems empower audiences to co-create narratives and influence reputation outcomes in real time. Scholars widely agree that the emergence of social platforms has redefined how organizations engage with stakeholders, recover from crises, and sustain brand reputation (Jin, Liu, & Austin, 2014; Fearn-Banks, 2016). In both the United States and Kazakhstan, digital communication now serves as both an opportunity and a risk, depending on how effectively organizations monitor, interpret, and respond to online discourse.

This section explores the relationship between digital media, public trust, and brand reputation, focusing on how organizations in both contexts utilize online communication tools to manage credibility and relational capital during crises.

7.1 The Rise of Participatory Communication

The shift toward participatory communication reflects a fundamental change in power dynamics between corporations and their audiences. In the digital environment, users can publicly challenge, reinterpret, and redistribute corporate messages. This phenomenon, often referred to as the democratization of communication, means that brand narratives are no longer unilaterally controlled (Jiang & Luo, 2018).

In the United States, corporations have adapted by incorporating real-time engagement strategies into their crisis management frameworks. For example, companies such as Starbucks and Delta Airlines have established dedicated social media response teams that address customer concerns within minutes. These teams rely on monitoring tools that track sentiment, identify misinformation, and adjust communication tone dynamically (Wigley & Zhang, 2017). The success of these strategies depends not only on speed but also on transparency, empathy, and consistency.

In Kazakhstan, the participatory turn is more recent but rapidly growing. Digital platforms such as Telegram, Instagram, and TikTok serve as spaces where citizens discuss

corporate behavior, share experiences, and express opinions. Companies like Kaspi.kz and Halyk Bank have begun using interactive posts, Q&A sessions, and short video statements to engage directly with their audiences. However, as Tulegenova and Sadykova (2022) note, the responsiveness of Kazakh corporations remains uneven. Some companies still treat digital platforms primarily as broadcasting tools rather than spaces for two-way dialogue. This transitional stage indicates that while participatory communication is emerging, it is not yet fully institutionalized in Kazakhstan's corporate landscape.

7.2 Social Media as a Trust-Building Mechanism

Trust is a core determinant of brand resilience. The Edelman Trust Barometer (2023) shows that 82 percent of consumers globally expect brands to act with honesty and transparency during crises. In digital spaces, trust is not only earned through words but also through visible engagement and responsiveness. According to Kim and Kim (2019), consistent and timely updates during a crisis enhance perceived competence and goodwill, leading to faster trust recovery.

U.S. corporations tend to leverage digital media to convey authenticity through humanized communication. For instance, during the 2020 Black Lives Matter protests, several brands issued direct social media statements acknowledging systemic issues and outlining internal reforms. Nike, Ben & Jerry's, and Netflix received positive public reactions due to the sincerity and specificity of their commitments (Carr & Hayes, 2021). These examples highlight that digital trust-building requires more than visibility; it demands value alignment between corporate messages and stakeholder expectations.

In Kazakhstan, digital trust-building efforts are at an earlier stage. Many corporations rely on formal press releases republished on social platforms, which limits emotional resonance. However, the public's increasing digital literacy is creating pressure for greater openness. For example, after the 2022 Almaty fuel protests, several Kazakh corporations used social media to express solidarity with citizens and clarify their operational safety policies. While some statements were criticized as performative, others were seen as a positive step toward transparency (Mukhamedova & Beknazarov, 2021). These developments suggest that trust-building in Kazakhstan is gradually evolving from institutional compliance toward relational engagement.

7.3 The Dual Nature of Digital Media: Opportunity and Risk

Digital media can both amplify and mitigate crises. On one hand, platforms enable direct communication, real-time corrections, and human-centered storytelling. On the other, they accelerate the spread of misinformation and outrage, often escalating minor incidents into large-scale reputational crises (Coombs & Holladay, 2018). This dual nature makes strategic digital literacy essential for modern crisis communicators.

In the United States, corporations have developed sophisticated monitoring and rapid-response mechanisms to manage this volatility. Social listening technologies, sentiment

analysis tools, and automated alerts allow communication teams to intervene early in misinformation cycles. For example, during the COVID-19 pandemic, companies such as Pfizer and Moderna utilized continuous social media engagement to clarify vaccine information and counter conspiracy narratives. These proactive strategies contributed to sustained public confidence in their corporate missions (Jin et al., 2021).

In Kazakhstan, the challenges are different. Although the digital landscape is vibrant, regulatory uncertainty and limited crisis training reduce companies' ability to navigate online controversies effectively. Tulegenova and Sadykova (2022) observe that some organizations respond to criticism by restricting comment sections or avoiding online discussions altogether. This approach may temporarily limit exposure but risks eroding long-term trust. To maximize the benefits of digital media, Kazakh corporations must invest in social media governance structures, media training, and real-time communication infrastructure.

7.4 Emotional Intelligence and Digital Tone

The tone of corporate communication on digital platforms is crucial in determining how audiences perceive sincerity and accountability. Jin, Liu, and Austin (2014) found that messages combining factual information with empathetic language generate higher trust recovery than purely informational content. Emotional intelligence, defined as the ability to recognize and respond appropriately to stakeholders' emotional states, is a growing focus in corporate communication research.

In the United States, companies often employ empathetic storytelling through social media. The use of personal narratives, behind-the-scenes videos, and employee testimonials humanizes corporate brands. Starbucks, for example, frequently integrates employee perspectives in its crisis responses to reinforce shared values of respect and inclusion (Fearn-Banks, 2016).

In Kazakhstan, emotional expression in corporate messaging remains less common. Communication tends to emphasize technical precision and institutional authority. This preference aligns with high-context cultural norms, where indirect communication and composure are valued (Hofstede, 2011). Nevertheless, some Kazakh corporations are beginning to adopt a more empathetic tone. For instance, Air Astana's CEO addressed passengers directly through a video statement following a safety incident in 2021, signaling a shift toward more human-centered crisis communication. This evolving tone suggests that emotional authenticity is gradually being recognized as a strategic component of reputation management in Kazakhstan.

7.5 Global Brand Reputation in the Digital Age

Brand reputation is now an interactive, data-driven construct that depends on collective perception across digital ecosystems. Fombrun and van Riel (2004) define reputation as a stakeholder's overall assessment of a company's reliability, quality, and integrity. In the digital age, this assessment is continuously shaped by user-generated content, influencer narratives, and algorithmic visibility.

In the United States, brand reputation management has evolved into a specialized discipline that integrates analytics, behavioral science, and strategic storytelling. Companies measure reputation through trust indices, media sentiment scores, and social media engagement metrics (Dowling, 2016). This data-driven approach allows for early identification of reputational risks and evaluation of recovery effectiveness.

In Kazakhstan, reputation management remains less formalized but increasingly influenced by international practices. Multinational corporations operating in the country have introduced reputation measurement tools and stakeholder surveys to assess trust and social impact (Mukhamedova & Beknazarov, 2021). Domestic companies are beginning to adopt similar approaches, although the emphasis often remains on public image rather than multidimensional stakeholder trust.

Ultimately, the digital environment is accelerating convergence between both contexts. The shared challenges of misinformation, transparency, and ethical branding are universal. Companies in both the U.S. and Kazakhstan must balance speed, authenticity, and accuracy to protect reputation in an environment of instantaneous global scrutiny.

8. Discussion and Critical Reflections

This section integrates theoretical insights and empirical observations from previous sections to provide a holistic understanding of how crisis communication functions as both a strategic and ethical process in the digital era. The comparative lens between the United States and Kazakhstan reveals how structural maturity, cultural context, and technological capability shape corporate responses to crises. The discussion extends beyond description by interpreting the implications of these findings for organizational practice, professional education, and future research in public relations and corporate communication.

8.1 Rethinking Crisis Communication in the Digital Context

The digital transformation of communication has fundamentally altered the crisis management landscape. The traditional linear model of information dissemination has been replaced by an interactive, multidirectional system where stakeholders play an active role in shaping brand narratives (Jin, Liu, & Austin, 2014). Organizations can no longer rely solely on centralized control; they must navigate a public sphere characterized by immediacy, transparency, and co-creation.

The cases analyzed in both contexts demonstrate that digital communication is both a strategic necessity and a potential vulnerability. In the United States, companies such as Nike and Boeing illustrate contrasting outcomes of digital engagement. Nike's proactive management of controversy through value-driven messaging strengthened its reputation, while Boeing's delayed acknowledgment of responsibility intensified public criticism. These examples reinforce the assertion by Coombs (2015) that strategic crisis communication must be guided by both the situational context and the organization's moral orientation.

In Kazakhstan, digitalization has accelerated public access to information and increased scrutiny of corporate behavior. Yet, the transition from reactive to preventive crisis communication remains incomplete. As observed in the cases of Air Astana and Kaspi.kz, Kazakh corporations are beginning to embrace digital responsiveness, but cultural and institutional constraints still limit the extent of open dialogue. This evolution illustrates that technological access alone does not guarantee communicative maturity; it must be supported by organizational learning and ethical leadership.

8.2 The Intersection of Culture and Strategy

Culture profoundly influences the perception of communication effectiveness. The comparative analysis reveals that while both U.S. and Kazakh corporations use similar communication frameworks, their application differs according to cultural values and audience expectations. In low-context cultures like the United States, directness, immediacy, and personal accountability are essential for maintaining credibility (Hofstede, 2011). In high-context cultures like Kazakhstan, communication tends to be more formal, cautious, and authority-oriented, which reflects broader social hierarchies and collectivist norms (Hall, 1976).

These cultural differences explain variations in tone, timing, and framing. American corporations often use public apology as a strategic tool for reputation repair, viewing transparency as a form of strength. In Kazakhstan, however, apologies may be interpreted as weakness or liability, leading companies to prioritize factual explanations and institutional authority. The challenge for Kazakh corporations lies in reconciling cultural norms with global expectations for openness and emotional authenticity.

Theories such as SCCT (Coombs, 2007) and SMCC (Jin et al., 2014) remain useful across contexts but require cultural adaptation. The success of crisis communication depends on the alignment between message strategy and audience expectations. As Taylor (2010) argues, international communication theories must be localized rather than simply transferred. This insight has practical implications for PR education in emerging markets, where training should emphasize contextual interpretation of global frameworks.

8.3 Ethical Responsibility as Strategic Advantage

Ethics is no longer an abstract concern in corporate communication; it is a determinant of long-term reputation and public legitimacy. Bowen (2008) asserts that ethical communication is the most sustainable form of reputation management because it aligns corporate action with stakeholder values. The U.S. context illustrates how companies that prioritize moral accountability can transform crises into opportunities for credibility reinforcement. Johnson & Johnson's handling of the Tylenol crisis remains a classic example of how transparency and empathy can rebuild trust even in the face of tragedy.

In contrast, the Kazakh experience highlights the early stages of ethical institutionalization. Many companies are still developing frameworks for social responsibility and ethical disclosure. As regulatory and consumer pressures increase, corporations in

Kazakhstan will need to move from reactive compliance toward proactive ethical communication. Global partnerships and corporate governance reforms can accelerate this process by embedding transparency and accountability into organizational culture (Mukhamedova & Beknazarov, 2021).

Ethics also intersects with digital communication. The speed and reach of social media amplify both truth and misinformation. Thus, ethical behavior in the digital space extends to fact verification, respectful dialogue, and responsible narrative framing. As Coombs and Holladay (2018) note, ethical crisis communication strengthens not only external trust but also internal cohesion, as employees view ethical leadership as a reflection of organizational integrity.

8.4 The Role of Leadership and Emotional Intelligence

Crisis communication effectiveness depends heavily on leadership behavior. Research consistently shows that CEOs and senior executives who communicate openly, empathetically, and consistently can significantly influence stakeholder trust (Fearn-Banks, 2016). In the United States, leadership visibility is an expected component of crisis response. Executives often appear in media briefings, personal statements, and live Q&A sessions. These actions convey accountability and humanize corporate institutions.

In Kazakhstan, leadership communication tends to be more reserved, often limited to official written statements. While this approach aligns with cultural respect for hierarchy, it can create perceptions of detachment. The gradual introduction of CEO-led communication, as seen in Air Astana's 2021 video address, represents progress toward international norms. Emotional intelligence, the capacity to perceive and manage emotions effectively, is increasingly recognized as a vital leadership skill in global corporate communication (Goleman, 2018). Leaders who demonstrate empathy and authenticity can mitigate reputational harm even before corrective actions are implemented.

The integration of emotional intelligence into corporate communication training could greatly benefit Kazakh companies. It would allow leaders to navigate crises with greater sensitivity and adaptability, ensuring that organizational responses resonate with diverse audiences. Moreover, fostering emotionally intelligent leadership supports a shift from reactive to relational crisis management, where communication is guided by long-term trust rather than short-term damage control.

8.5 Implications for Public Relations Education and Professionalization

The comparative findings of this study suggest that effective crisis communication requires a combination of theoretical knowledge, cultural competence, and technological literacy. Public relations education must evolve to meet these interdisciplinary demands. In the United States, academic programs increasingly incorporate case-based learning, simulation exercises, and digital analytics to prepare students for the complexities of crisis communication (Heath, 2010).

In Kazakhstan, PR education remains largely theoretical, with limited emphasis on applied crisis management or cross-cultural communication. Strengthening university curricula through partnerships with international institutions could address this gap. Exposure to global best practices would enable future Kazakh communication professionals to integrate international standards while preserving local relevance. Training programs should include modules on digital media ethics, stakeholder engagement, and intercultural negotiation to prepare graduates for global communication roles.

Professional associations such as the Public Relations Society of America (PRSA) and the European Public Relations Education and Research Association (EUPRERA) provide frameworks that could be adapted regionally to support professional certification and ethical standardization. As the Kazakh communication sector matures, formalized training and accreditation systems will be essential to institutionalize best practices and ensure accountability.

8.6 The Globalization of Reputation and the Path Forward

The globalization of communication has blurred national boundaries, making reputational crises instantly visible worldwide. Both U.S. and Kazakh corporations operate within a shared digital ecosystem where stakeholder perceptions transcend geography. This reality underscores the need for integrated global-local strategies that combine universal principles of transparency and accountability with culturally appropriate communication techniques.

Future crisis management will depend increasingly on predictive analytics, artificial intelligence, and big data monitoring to anticipate reputational risks before they escalate. However, technology cannot replace the human elements of trust, empathy, and ethical judgment. The future of crisis communication lies in balancing technological precision with emotional intelligence and moral responsibility.

Kazakhstan's evolving communication environment represents a valuable laboratory for studying the localization of global frameworks. As its economy diversifies and public discourse becomes more open, the country will likely serve as a regional example of how emerging markets can modernize corporate communication without eroding cultural identity. Conversely, U.S. corporations may draw lessons from Kazakhstan's adaptive resilience and context-sensitive approach to communication, reminding practitioners that global best practices must always respect local nuance.

9. Future Directions and Research Implications

Crisis communication continues to evolve at the intersection of culture, technology, and public accountability. The comparative analysis of the United States and Kazakhstan reveals not only the complexity of managing crises in diverse contexts but also the opportunities for global learning and institutional innovation. As corporate environments become more digitized and interconnected, future research must expand its scope to include emerging technologies, evolving public expectations, and interdisciplinary approaches to understanding reputation, trust, and communication ethics.

9.1 Integrating Technology and Predictive Analytics

One of the most significant future directions for crisis communication research lies in the integration of predictive analytics and artificial intelligence (AI). The growing use of machine learning algorithms for media monitoring, sentiment analysis, and early warning systems is reshaping how organizations identify and respond to potential crises (Liu et al., 2018). Predictive models can analyze large datasets to detect anomalies in brand perception, enabling companies to intervene before a situation escalates.

In the United States, predictive analytics is already being embedded into strategic communication frameworks. For instance, major corporations use AI tools to assess public sentiment and simulate possible response outcomes (Wigley & Zhang, 2017). In Kazakhstan, this trend is still nascent but holds significant potential. As digital infrastructure improves, the adoption of AI-based monitoring tools could enhance the agility and responsiveness of corporate crisis teams.

However, researchers and practitioners must also address the ethical and technical limitations of predictive technologies. Issues of data privacy, algorithmic bias, and overreliance on automation can undermine credibility and accountability. Future studies should explore how organizations can balance technological innovation with ethical transparency, ensuring that AI complements rather than replaces human judgment in communication decision-making.

9.2 Expanding Cross-Cultural Comparative Research

Crisis communication scholarship remains heavily concentrated in Western contexts. Although models such as SCCT and SMCC have proven useful globally, their universal applicability remains limited without robust cross-cultural validation. Future research should prioritize comparative studies across diverse regions, particularly in emerging economies like Kazakhstan, where institutional and cultural dynamics differ from Western assumptions (Taylor, 2010).

Longitudinal studies could provide insights into how globalization, media liberalization, and generational change influence communication behavior over time. For example, as younger, digitally fluent generations become corporate decision-makers in Kazakhstan, communication strategies may align more closely with global best practices. Quantitative cross-national research that measures public trust, transparency perception, and response effectiveness could provide empirical grounding for theoretical refinements.

Additionally, future research should examine hybrid communication models that combine local traditions of interpersonal trust and collective responsibility with modern principles of openness and engagement. Such models would contribute to the development of a truly global theory of crisis communication that respects cultural diversity while promoting ethical universality.

9.3 Professionalization and Policy Development

The professionalization of public relations and corporate communication remains uneven globally. The U.S. benefits from well-established professional organizations, academic programs, and ethical codes such as those outlined by the Public Relations Society of America (PRSA, 2022). These structures ensure consistent standards and accountability across industries. In Kazakhstan, professionalization is advancing but still requires institutional reinforcement. Developing a national PR association, standardized certification programs, and ethics committees could significantly strengthen the field. Partnerships between Kazakh universities and international communication associations could facilitate curriculum development, research collaboration, and practitioner exchange.

From a policy perspective, governments can play an enabling role by supporting transparency regulations and corporate disclosure requirements. The introduction of accountability frameworks for environmental, social, and governance (ESG) reporting has already encouraged many Kazakh corporations to enhance communication transparency (Mukhamedova & Beknazarov, 2021). Continued integration of communication ethics into regulatory policy would further institutionalize best practices and strengthen public trust in both corporate and governmental institutions.

9.4 Interdisciplinary Integration

Future research should also move toward interdisciplinary collaboration. Crisis communication does not exist in isolation; it intersects with fields such as behavioral psychology, data science, organizational ethics, and information technology. Integrating these disciplines would enable scholars and practitioners to develop multidimensional models of crisis perception and response.

For instance, combining neuroscience and psychology can shed light on emotional processing and public empathy during crises, while data science can enhance precision in sentiment measurement and stakeholder mapping (Goleman, 2018). Similarly, political science perspectives can illuminate how institutional trust influences corporate legitimacy. By bridging these disciplinary divides, researchers can design comprehensive frameworks that capture the cognitive, emotional, and structural dimensions of crisis communication.

9.5 Education, Training, and Public Literacy

Finally, the future of crisis communication depends on education and public literacy. Organizations must invest not only in professional training but also in educating the general public about digital responsibility and information verification. Public understanding of credible sources, misinformation, and ethical engagement can significantly reduce the social amplification of crises.

Universities and communication programs, particularly in emerging economies like Kazakhstan, should integrate practical training on digital ethics, cross-cultural negotiation, and stakeholder psychology. Experiential learning methods such as simulation-based crisis exercises and cross-border collaborative projects can prepare students to manage communication

challenges in a globalized world. The development of public awareness campaigns about media literacy and corporate accountability can also build a more resilient information environment. These efforts require cooperation between the public sector, academia, and private industry to foster a communication culture rooted in trust, ethics, and critical thinking.

10. Conclusion

This review has examined the evolution of crisis communication and brand reputation management in the digital era, comparing corporate response strategies in the United States and Kazakhstan. The analysis demonstrates that while global frameworks such as Situational Crisis Communication Theory (SCCT) and the Social-Mediated Crisis Communication (SMCC) Model provide universal foundations, their application remains deeply influenced by institutional maturity, cultural values, and technological infrastructure.

The United States represents a mature communication environment characterized by institutionalized ethical norms, rapid digital engagement, and strategic integration of crisis management into corporate governance. American corporations such as Johnson & Johnson, Nike, and Boeing exemplify both the successes and failures that define best practices in crisis response. In contrast, Kazakhstan's crisis communication field is in a formative phase, gradually transitioning from reactive information control to proactive stakeholder dialogue. While cultural norms emphasizing hierarchy and social harmony still shape corporate messaging, increasing digitalization and international exposure are fostering transparency and engagement.

The comparative analysis underscores three critical insights. First, digital communication has universalized the expectation of immediacy and authenticity, compelling organizations across cultures to respond in real time. Second, ethics and empathy remain central to public trust and cannot be replaced by technology or efficiency. Third, effective crisis communication depends on cross-cultural adaptability. Organizations must align global standards with local sensibilities to sustain credibility in diverse markets.

As both countries continue to navigate the challenges of digital transformation, the convergence of technological innovation, ethical leadership, and cultural literacy will define the future of global corporate communication. The lessons from this comparative study offer guidance not only for practitioners and scholars but also for policymakers seeking to strengthen institutional transparency and stakeholder trust. Ultimately, the practice of crisis communication must evolve from a reactive defensive function into a proactive instrument of public accountability and corporate integrity.

Works Cited

- Austin, L., & Jin, Y. (2017). Social media and crisis communication: Expanding the boundaries of social-mediated crisis communication. *Public Relations Review*, 43(3), 593–599. <https://doi.org/10.1016/j.pubrev.2017.02.014>
- Benoit, W. L. (1995). *Accounts, excuses, and apologies: A theory of image restoration strategies*. State University of New York Press.
- Bowen, S. A. (2008). A state of neglect: Public relations as “corporate conscience” or ethics counsel. *Journal of Public Relations Research*, 20(3), 271–296. <https://doi.org/10.1080/10627260801962749>
- Carr, C. T., & Hayes, R. A. (2021). Social media and the moral economy of brand activism: A critical review. *Journal of Communication Management*, 25(4), 329–345. <https://doi.org/10.1108/JCOM-06-2020-0056>
- Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of situational crisis communication theory. *Corporate Reputation Review*, 10(3), 163–176. <https://doi.org/10.1057/palgrave.crr.1550049>
- Coombs, W. T. (2014). *Ongoing crisis communication: Planning, managing, and responding* (4th ed.). Sage Publications.
- Coombs, W. T. (2015). The value of communication during a crisis: Insights from strategic communication research. *Business Horizons*, 58(2), 141–148. <https://doi.org/10.1016/j.bushor.2014.10.003>
- Coombs, W. T., & Holladay, S. J. (2012). The paracrisis: The challenges created by publicly managing crisis prevention. *Public Relations Review*, 38(3), 408–415. <https://doi.org/10.1016/j.pubrev.2012.04.004>
- Coombs, W. T., & Holladay, S. J. (2018). *The handbook of crisis communication*. Wiley-Blackwell.
- Dowling, G. R. (2016). Reputation risk: It’s the board’s ultimate responsibility. *Journal of Business Strategy*, 37(1), 41–49. <https://doi.org/10.1108/JBS-09-2015-0090>
- Edelman. (2023). *Edelman Trust Barometer 2023: Navigating a polarized world*. Edelman. <https://www.edelman.com/trust/2023-trust-barometer>
- Elliott, D., & Hanna, S. (2020). Reassessing crisis communication response: Boeing and the 737 MAX. *Public Relations Inquiry*, 9(3), 245–260. <https://doi.org/10.1177/2046147X20917639>
- Fearn-Banks, K. (2016). *Crisis communications: A casebook approach* (5th ed.). Routledge.
- Fombrun, C. J., & van Riel, C. B. M. (2004). *Fame and fortune: How successful companies build winning reputations*. Pearson Education.
- Gillespie, N., & Dietz, G. (2009). Trust repair after an organization-level failure. *Academy of Management Review*, 34(1), 127–145. <https://doi.org/10.5465/amr.2009.35713319>
- Goleman, D. (2018). *Emotional intelligence: Why it can matter more than IQ*. Bantam Books.
- Hall, E. T. (1976). *Beyond culture*. Anchor Books.
- Heath, R. L. (2010). *The Sage handbook of public relations*. Sage Publications.

- Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture*, 2(1), 1–26. <https://doi.org/10.9707/2307-0919.1014>
- Huang, Y. H. (2000). The personal influence model and Gao Guanxi in Taiwan Chinese public relations. *Public Relations Review*, 26(2), 219–236. [https://doi.org/10.1016/S0363-8111\(00\)00039-1](https://doi.org/10.1016/S0363-8111(00)00039-1)
- Jiang, H., & Luo, Y. (2018). Crafting crisis communication strategies in the digital era. *Corporate Communications: An International Journal*, 23(2), 320–336. <https://doi.org/10.1108/CCIJ-11-2016-0073>
- Jin, Y., Liu, B. F., & Austin, L. L. (2014). Examining the role of social media in effective crisis management: The effects of crisis origin, information form, and source on public responses. *Communication Research*, 41(1), 74–94. <https://doi.org/10.1177/0093650211423918>
- Kantar Central Asia. (2022). Digital media consumption trends in Kazakhstan 2022. Kantar Media Insights.
- Kim, H. J., Avery, E. J., & Lariscy, R. W. (2009). Are crisis communicators practicing what we preach? An evaluation of crisis response strategy analyzed by situational crisis communication theory. *Public Relations Review*, 35(4), 446–448. <https://doi.org/10.1016/j.pubrev.2009.08.002>
- Kim, S., & Kim, J. (2019). The impact of corporate social responsibility communication on public trust. *Public Relations Review*, 45(3), 101–119. <https://doi.org/10.1016/j.pubrev.2019.03.005>
- Makhmutova, D., & Sadykova, G. (2020). Corporate communication culture in Kazakhstan: An emerging field of practice. *Central Asian Journal of Communication Studies*, 3(2), 45–63.
- Mukhamedova, S., & Beknazarov, T. (2021). Ethical communication and CSR in Central Asia: Emerging trends and challenges. *Journal of Business Ethics and Communication*, 6(1), 88–103.
- Parganas, P., & Anagnostopoulos, C. (2020). Brand activism in sportswear: Lessons from Nike's Kaepernick campaign. *Journal of Brand Management*, 27(6), 673–688. <https://doi.org/10.1057/s41262-020-00213-2>
- Public Relations Society of America (PRSA). (2022). Code of Ethics. PRSA. <https://www.prsa.org/ethics>
- Shin, S., Cho, M., & Cheng, I. (2021). Learning from Boeing's communication mistakes: The impact of crisis response strategies on reputation. *Corporate Reputation Review*, 24(1), 45–60. <https://doi.org/10.1057/s41299-020-00102-z>
- Sriramesh, K., & Vercic, D. (2019). *The global public relations handbook: Theory, research, and practice* (3rd ed.). Routledge.
- Taylor, M. (2010). Toward a global public relations theory. *Public Relations Inquiry*, 1(2), 195–210. <https://doi.org/10.1177/2046147X10369228>
- Tulegenova, S., & Sadykova, G. (2022). The evolution of public relations and corporate communication in Kazakhstan. *Eurasian Journal of Social Sciences and Humanities*, 9(1), 33–54.
- Wigley, S., & Zhang, W. (2017). Tools of engagement: Social media management and crisis response. *Journal of Communication Management*, 21(3), 270–289. <https://doi.org/10.1108/JCOM-06-2016-0044>

Microglial Activation in Children with Early Cortical Thinning Predicts ADHD Symptoms: Attention, Memory, and Planning Deficits in Adolescence

By Vishika Challapalli

Abstract

Neuroinflammation driven by microglia, the brain's primary immune cells, is increasingly recognized as a key factor in both neurodegenerative and neurodevelopmental disorders. However, the role of microglia in early brain development and executive dysfunction in children, including those with Attention-Deficit/Hyperactivity Disorder (ADHD), remains less understood. The present literature review identifies key studies examining the relationship between microglial activation, cortical structure, and executive function. A positron emission tomography (PET) imaging study of microglial activation in adults with frontotemporal dementia found that early microglial activation predicted subsequent cognitive decline independently of visible brain atrophy. Pediatric magnetic resonance imaging (MRI) studies on cortical thinning and executive dysfunction in children with ADHD linked cortical thinning in prefrontal areas with impaired executive skills and ADHD symptoms. Finally, executive behavior studies in mice exploring the effects of microglial depletion on brain development demonstrated that microglial depletion during adolescence disrupts normal synaptic pruning and dendritic maturation, leading to long-lasting deficits in cognitive flexibility and executive functioning. The converging evidence reveals that increased microglial activation in the frontal cortex precedes or accompanies cortical thinning and is strongly associated with deficits in executive functions such as attention, working memory, and planning. This integrated framework highlights the promise of microglial activity as an early biomarker and therapeutic target for preserving cognitive function in at-risk youth.

Introduction

Inflammation in the brain, or neuroinflammation, refers to the sustained immune response within the central nervous system. In a healthy body, inflammation is a response to foreign invasion by pathogens or injury, such as stroke or traumatic brain injury. The immune response can be assigned to destroy pathogens and infected cells, create barriers between damaged tissue and healthy tissue, stop bleeding, reduce energy output, and recruit other immune cells to the site of injury. In all cases, the immune response alters the physiology of nearby cells, which can be extremely damaging if prolonged inflammation persists.

Microglia are the brain's primary immune cells. They constantly survey the brain environment and respond to injury or infection. When activated, they release inflammatory molecules, such as cytokines IL-1 β and TNF- α . In high concentrations or when present for long periods of time, these cytokines can impair synaptic functioning and disrupt neural circuits (DiSabato et al. 2016). The impact of inflammation on the brain is of special interest in aging and age-linked diseases like stroke and dementia, where inflammation in these diseases has long

been observed. However, excess inflammation in childhood may contribute to childhood diseases as well.

Inflammation plays a critical role in the progression of dementia, where chronic activation of the brain's immune system, particularly microglia, contributes to synaptic loss and neuronal dysfunction. In healthy brains, microglia are essential for synapse pruning, remodeling, and maintenance, all of which help to shape neural circuits during development and maintain cognitive function in adulthood. However, when chronically activated, as in neuroinflammatory states, microglia can become neurotoxic, releasing cytokines and reactive oxygen species (ROS) that disrupt synaptic signaling and plasticity. Given their central role in synapse development, dysregulated microglia during early brain development could have far-reaching consequences.

This microglial dysregulation is especially significant in the context of neurodevelopment, where the balance between synaptic pruning and formation is tightly regulated. If microglial function is altered during critical periods of cortical maturation, it may lead to atypical neural connectivity, contributing to neurodevelopmental disorders, such as ADHD and autism spectrum disorder. Recent studies suggest that early cortical thinning, potentially driven by abnormal microglial activation, is associated with attentional deficits and poor executive function outcomes. Furthermore, emerging evidence also links early-life neuroinflammation with mood disorders like depression, indicating that microglia may play a foundational role in the trajectory of mental health. Understanding microglial contributions to early brain development could therefore provide key insights into the mechanisms underlying ADHD and other cognitive or affective disorders. (Narr et al. 2009; McLaughlin et al. 2014).

Neurodevelopmental Conditions

Attention-Deficit/Hyperactivity Disorder (ADHD) is a neurodevelopmental condition characterized by persistent patterns of inattention, hyperactivity, and impulsivity that interfere with daily functioning or development. It is associated with atypical development of brain regions involved in executive function, particularly the prefrontal cortex, which is responsible for attention regulation, impulse control, and working memory. Neuroimaging studies have shown delayed cortical maturation in children with ADHD, especially in frontal and parietal regions (Shaw et al. 2007). Dysregulation of dopamine and norepinephrine signaling in frontostriatal circuits also contributes to difficulties in sustaining attention and controlling behavior (Arnsten, 2009). Although symptoms often emerge in early childhood, ADHD can persist into adolescence and adulthood, with impacts on academic performance, occupational success, and emotional regulation. Early intervention and multimodal treatment, including behavioral therapy and medication, can significantly improve outcomes.

Vulnerability of the Prefrontal Cortex

The prefrontal cortex, which plays a central role in planning, decision-making, memory, and sustained attention, is one of the last brain regions to mature fully.

Studies involving healthy children have found that the thickness of the prefrontal cortex is predictive of performance on cognitive tasks. In healthy children, it plays a key role in regulating behavior and supporting goal-directed actions. In children with ADHD, this region often functions abnormally, leading to difficulties with focus, organization, and self-regulation. The underlying causes of this dysfunction are not fully understood but believed to involve both structural and functional brain differences. Factors such as altered neural connectivity, neurotransmitter imbalances, and early developmental disruptions are thought to contribute.

Rationale for the Current Study

This study aims to examine these clinical, developmental, and mechanistic findings to investigate whether microglial activation in the frontal cortex acts as an early biomarker and possible driver of disrupted brain development. A greater understanding of this mechanism—the neuroinflammation and consequent cortical thinning leading to impairments in attention, planning, and memory in children—could open the door to early detection and intervention strategies aimed at preserving executive function during critical developmental periods.

Results

Cortical Thinning and Dementia

Cortical thinning is a well-established neuroimaging marker of early-stage dementia, particularly in Alzheimer's disease (AD). In a seminal study, Dickerson et al. demonstrated that thinning in specific cortical areas—such as the medial temporal lobe, posterior cingulate, and lateral prefrontal cortex—was significantly associated with the progression from mild cognitive impairment (MCI) to AD, with areas of the temporal lobe having the strongest correlation between thinning and AD. Importantly, this cortical thinning was predictive of cognitive decline even before clinical symptoms became pronounced, supporting its utility as an early biomarker. The study also found that these patterns of thinning correlated with symptom severity in very mild to mild AD, further confirming the link between structural changes and functional outcomes (Dickerson et al. 2009). These findings suggest that cortical thinning is not only a consequence of neurodegeneration but may also serve as a measurable indicator of future cognitive deterioration.

Effect of Microglial Activation on Cognitive and Mental Outcomes Overtime

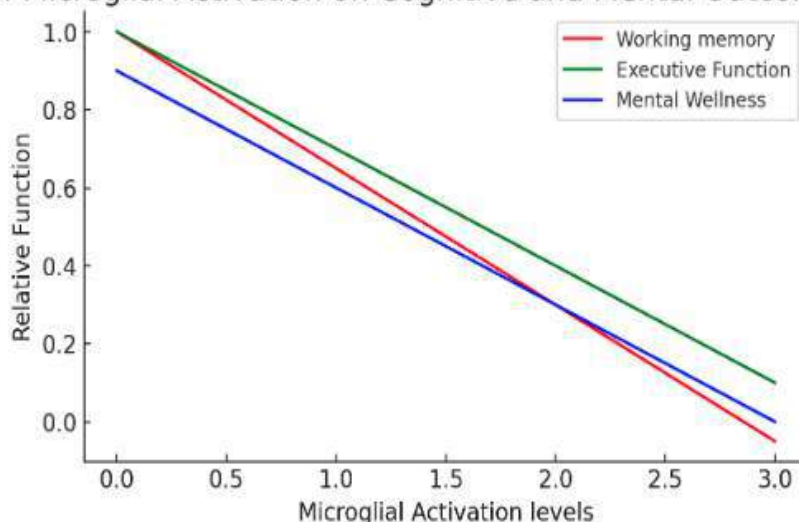


Figure 1. Schematic illustrating the relationship between microglial activation and cognitive decline over time. This figure demonstrates the impact of microglial activation on three key domains of cognitive and psychological function: working memory, decision-making, and mental health. Elevated levels of microglial activation, particularly during critical developmental periods such as adolescence, are associated with progressive decline in these areas. Working memory is the most affected, as microglial overactivation disrupts synaptic pruning in the prefrontal cortex, leading to long-term impairments in the ability to hold and manipulate information. This function is essential for learning and goal-directed behavior. In decision-making, abnormal microglial activity interferes with the maturation of circuits in the medial prefrontal cortex, resulting in difficulties evaluating choices, predicting outcomes, and regulating impulsive behavior. Mental health is also impacted, as chronic microglial activation contributes to neuroinflammation, which has been linked to increased vulnerability to anxiety and depression. These effects likely result from altered neural plasticity and dysregulated stress-response pathways during development. The data and concepts illustrated in this figure are adapted from Schalbetter et al. and DiSabato et al.

Cortical Thinning is Correlated with Microglia Activation

Determining the role that elevated microglial activation in the frontal lobe plays in predicting cognitive decline in children with early cortical thinning requires synthesizing evidence across clinical, neuroimaging, and animal model studies. Foundational studies examining cortical thinning and cognitive decline were performed on adult patients with frontotemporal dementia (FTD), where degeneration of the frontal cortex and cognitive decline are pronounced. Bevan-Jones et al. utilized positron emission tomography (PET) imaging with translocator protein (TSPO) tracers, which selectively bind to activated microglia, to measure neuroinflammation in the frontal cortex of FTD patients (2020). Their longitudinal study found that individuals exhibiting higher baseline frontal microglial activation experienced accelerated cognitive decline in domains critical for executive functioning, including attention, planning, and memory.

Crucially, these associations persisted even when accounting for the severity of cortical atrophy at the beginning of the study, suggesting that neuroinflammation does not merely follow structural degeneration but may actively drive cognitive deterioration. This is a critical insight, as

it highlights microglial activation as a potential independent biomarker for executive dysfunction progression, rather than a secondary consequence of neuronal loss.

Additionally, this study highlights that neuroinflammation and aggregation of misfolded proteins co-localize in the frontal regions, providing a neuropathological substrate for microglial-mediated damage. Protein aggregation in FTD acts as the upstream insult that activates microglia. The resulting neuroinflammation accelerates cortical thinning, disrupting frontal and temporal circuits that underpin executive function, emotional control, and social behavior. This supports the hypothesis that inflammation, reflected in microglial activity, disrupts neural circuits governing executive processes. Thus, elevated microglial activation in adults with neurodegeneration establishes a precedent for exploring its predictive value in earlier, pediatric populations where cortical thinning is present but neurodegeneration is not yet severe.

Table 1. Data table of cortical thickness measured in millimeters (mm) in the temporal lobe in relation to cognitive decline of participants with Alzheimer's disease with a control group

	OC (Older nondemented controls)	Very mild AD (Alzheimer's disease)	Difference (mm)
Temporal Pole	2.63	2.52	0.11
Medial Temporal Lobe	2.91	2.63	0.28
Inferior Temporal Lobe	2.67	2.53	0.14
$\bar{x}$	2.736	2.56	

According to findings presented by Dickerson et al. (2009), measurements of cortical thickness in the temporal lobe reveal a significant correlation between cortical atrophy and cognitive decline, particularly in individuals with Alzheimer's disease (AD). The data illustrate a clear contrast between older nondemented individuals (control group) and those diagnosed with AD. The control group exhibited relatively preserved cortical thickness in key regions of the temporal lobe, which is consistent with maintained cognitive function, while participants with AD demonstrated reduced cortical thickness in these same regions. Data from Dickerson et al. 2009.

Cortical Thinning Impairs Academic and Social Success

Is cortical thinning associated only with dementia, or does it occur in other disorders? While it is a hallmark feature of neurodegenerative conditions like Alzheimer's disease, cortical thinning is increasingly recognized as a transdiagnostic biomarker found across a range of

psychiatric and neurodevelopmental disorders. Schizophrenia, depression, and autism spectrum disorder (ASD), for instance, all show patterns of cortical thinning in frontal and temporal regions, often accompanied by deficits in cognition, emotion regulation, or social processing. These findings suggest that cortical thinning, though prominent in aging-related pathology, is not exclusive to late-life conditions but is also deeply involved in atypical brain development.

In pediatric populations, studies have increasingly demonstrated that cortical thinning, particularly in frontal and temporal regions, correlates strongly with deficits in executive function. These regions are critical for regulating attention, working memory, and planning—functions that are commonly impaired in children with neurodevelopmental disorders. Cortical thickness in these areas typically follows a nonlinear developmental trajectory, with peak thickness in childhood followed by gradual thinning into adolescence, caused by strategic synaptic pruning and increased neural efficiency. Deviations from this trajectory, such as early or excessive thinning, may therefore signal disruptions in the maturation of executive systems.

To investigate the role of cortical thickness in younger individuals, Narr et al. conducted a study of cortical thickness in children with ADHD and similarly aged controls (2009). In children diagnosed with ADHD, Narr et al. found widespread reductions in cortical thickness, especially within the prefrontal cortex, that aligned with impairments in working memory, attention, and planning (2009). The prefrontal cortex, responsible for regulating goal-directed behavior and inhibiting inappropriate responses, is a central node in the executive function network. Thinning in this region may reflect abnormal neurodevelopmental processes, including altered synaptic pruning and myelination, that ultimately compromise cognitive control. These structural differences may underlie the behavioral symptoms of ADHD, suggesting that the disorder may involve both delayed and disrupted cortical maturation.

These findings raise a broader question: is atypical cortical thinning specific to ADHD, or does it represent a more general pathway by which diverse developmental risks impair executive functioning? To address this question, McLaughlin et al. studied children who have experienced environments with a severe lack of expected social, emotional, and cognitive inputs during early development, as these experiences are implicated in executive dysfunction later in life. The study found significant cortical thinning in frontal areas that predicted the severity of executive dysfunction (2014). Additionally, these findings highlight the profound impact of environmental stressors on brain development through behavioral assays. Children raised in environments lacking cognitive and emotional stimulation often exhibit accelerated or abnormal thinning in areas critical for self-regulation and planning, indicating that cortical thickness may be a link between childhood trauma and later executive dysfunction. In McLaughlin et al.'s study (2014), these environments typically involved institutional rearing, where children experienced minimal caregiver interaction, limited language exposure, and little opportunity for play or emotional support. Such deprivation deprived the brain of expected inputs needed to drive healthy cortical maturation. As a result, children exposed to these conditions showed reductions in cortical thickness of about 0.1–0.2 mm in frontal regions compared to typically developing peers, a meaningful difference given that average cortical thickness in children is about 2–4 mm. The

degree of thinning was directly related to poorer performance on executive function tasks such as working memory and inhibitory control, reinforcing that these structural differences had significant behavioral consequences (see Figure 2). This work underscores environmental impacts on neurodevelopmental trajectories and establishes cortical thickness as a meaningful biomarker tied to cognitive outcomes. It also suggests that early interventions aimed at reducing stress and enriching environments may help mitigate long-term executive impairments of at-risk children.

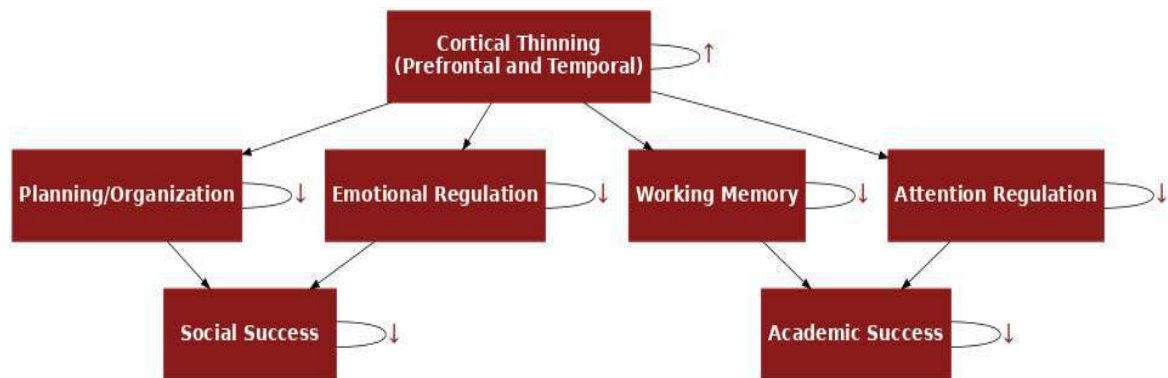


Figure 2. Diagram of connections of cortical thinning and its impact on social and academic success. This schematic illustrates how atypical cortical thinning in frontal and temporal regions disrupts the development of executive systems and contributes to functional impairments in pediatric populations. Cortical thinning, particularly when premature or excessive, compromises executive abilities such as working memory, attention control, planning, and emotional regulation. These deficits, in turn, directly affect children’s academic success (e.g., reduced learning efficiency, lower test performance, difficulty completing assignments) and social success (e.g., impaired peer relationships, poor conflict resolution, reduced acceptance in social groups). The figure highlights the cascading relationship between brain structure, cognitive function, and real-world outcomes, underscoring cortical thickness as a clinically meaningful biomarker for academic and social development. This schematic illustrates that abnormally high levels of cortical thinning result in lower social and academic success through a lack of planning/organization, emotional regulation, working memory, and attention regulation skills (Shaw et al. 2011).

Building on these mechanistic insights, developmental neuroimaging studies provide further evidence that the timing of cortical maturation is as critical as cortical thickness. In a longitudinal study of typically developing children and children with ADHD, Shaw et al. illustrated the importance of developmental timing, as represented by Figure 3 (2011). Their longitudinal work demonstrated that in typically developing children, the prefrontal cortex reaches maximal thickness at approximately 7–8 years of age, after which it thins gradually. It appears that cortical thinning in adolescence is vital for network refinement and cognitive efficiency. Children with ADHD also typically exhibit times of cortical growth and thinning. However, in typically developing children, cortical thickness peaks earlier (around 7–8 years) and thins gradually for approximately 10 years. In children with ADHD, however, the peak is delayed, occurring at approximately 10–11 years of age, indicating a lag in cortical maturation. In fact, more delayed peak cortical thickness in prefrontal regions correlated with more severe symptoms of hyperactivity and impulsivity (Shaw et al. 2011). Although the timing of the peak

differed, the actual maximum cortical thickness was similar between groups, and both groups ultimately converged toward comparable adult thickness. Thus, the key difference lay in the trajectory of cortical maturation rather than the final thickness itself. Was the end thickness different? These findings support and suggest that deviations in cortical maturation timing are directly linked to executive deficits. Rather than being static, brain development in ADHD appears to follow an altered trajectory, which may help explain why some symptoms improve with age while others persist.

Cortical Thinning In Participants With ADHD vs. Typically Developing Participants

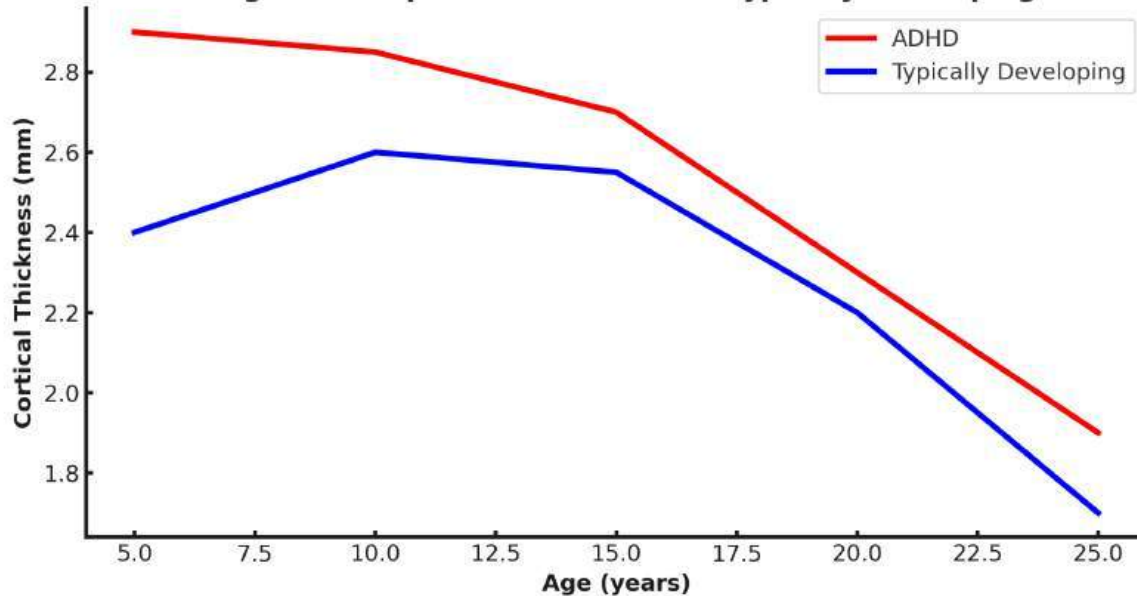


Figure 3. Graph illustrating the level of cortical thickness measured in millimeters in participants with ADHD, as well as participants who are developing typically with no other factors, based on their age measured in years. This figure shows developmental trajectories of cortical thickness in children with ADHD compared to typically developing peers. In typically developing participants, cortical thickness is greater in early childhood, peaks earlier, and then undergoes gradual thinning across adolescence, reflecting normal synaptic pruning and neural specialization. In contrast, children with ADHD display a delayed and reduced cortical thickness trajectory, with overall thinner cortices and a later peak. This deviation suggests altered cortical maturation, particularly in the prefrontal regions, which is linked to executive function deficits such as impaired attention, working memory, and planning (Shaw et al. 2011).

In addition to these clinical cohorts, normative developmental studies highlight that healthy thinning of the prefrontal cortex during adolescence supports maturation of planning and working memory capabilities. For example, studies have shown that cortical thickness in the right medial frontal gyrus predicted planning performance in healthy children and adolescents, suggesting that proper cortical development underpins executive ability (Luciana et al. 2012). Specifically, healthy children with higher cortical thickness were able to plan and organize problem-solving strategies more effectively when their cortical thickness was greater. This pattern of thinning is believed to reflect optimal synaptic pruning, where unnecessary neural

connections are eliminated to increase processing efficiency. These normative data offer a baseline against which atypical patterns can be identified and interpreted.

Together, these pediatric studies confirm that early cortical thinning, defined as an atypically rapid or premature reduction in cortical thickness from early childhood into adolescence, is not a benign process but is linked with functional impairments in executive domains critical for academic and social success. When thinning occurs prematurely, excessively, or in regions undergoing critical developmental change, it may disrupt the formation of networks essential for attention, memory, and behavioral control. This creates a compelling rationale for investigating upstream factors like microglial activation that may drive or exacerbate thinning.

Microglia are Necessary for Executive Function in Mice

An important part of cortical maturation is the selective growth and pruning of synapses; this pruning is especially important in cortical thinning. Microglia notably play a key role in synaptic pruning and respond dynamically to environmental signals, inflammation, and stress. Dysregulation of microglial function during development could therefore disturb cortical maturation, laying the groundwork for the emergence of ADHD symptoms and related executive dysfunction. In animal models, recent research has supported the idea that microglia play a causal role in shaping the development of executive function. Parkhurst et al. (2022) conducted a key study in which microglia were temporarily depleted from the prefrontal cortex of adolescent mice using localized drug injections that eliminated roughly 70–80% of microglia without harming surrounding brain cells. Compared to control animals that received a placebo or sham surgery, these mice showed profound and long-lasting deficits in planning, working memory, and cognitive flexibility, impairments that persisted well into adulthood. Structural analyses revealed that these cognitive deficits were accompanied by abnormalities in synaptic pruning and dendritic spine maturation, processes that microglia normally regulate through direct contact with neurons and the release of growth-promoting molecules. By interfering with these functions, microglial depletion disrupted the refinement of neural circuits critical for higher-order cognition. Notably, when the same procedure was performed in adult mice, no comparable impairments were observed, underscoring that adolescence represents a sensitive period during which microglial activity is indispensable for prefrontal cortical development and cognitive maturation.

To evaluate the long-term effects of this microglial loss, the researchers later tested the mice in adulthood using a range of behavioral tasks. These included assessments of memory for the order of objects, social interaction, the ability to unlearn fear associations, and several tasks focused on planning and flexibility in thinking. Mice that had experienced microglial depletion during adolescence performed significantly worse on these tests compared to control animals (Schalbetter et al. 2022). Importantly, when the microglial depletion was carried out in adult mice rather than adolescents, the same cognitive problems did not appear, suggesting that adolescence is a uniquely sensitive window for brain development (2022).

These results showed that microglial depletion disrupted synaptic pruning and dendritic spine maturation. Synaptic pruning is an important developmental process where extraneous or weak connections between neurons are removed to make brain networks more efficient and precise. At a larger scale, this process contributes to cortical thinning during childhood and adolescence, which reflects the brain becoming more specialized. This streamlining allows the brain to work more effectively, especially in regions involved in higher-level thinking. Dendritic spine maturation refers to the strengthening and stabilization of the small branches on neurons where connections form; mature spines support stronger and more reliable communication between neurons. Microglia play an active role in these processes by interacting with dendritic spines, releasing growth-supporting molecules, and responding to signals from neurons. In doing so, they help strengthen useful connections while removing weaker ones. This back-and-forth communication between microglia and neurons ensures that brain networks become more efficient and specialized as development progresses.

However, inflammatory signals can significantly disrupt microglia's ability to perform these critical functions. When microglia are exposed to pro-inflammatory cytokines or other immune challenges, they shift into a reactive state characterized by altered morphology and impaired signaling capacity. In this activated state, microglia may excessively prune synapses or fail to support proper spine maturation, leading to either a loss of necessary connections or the persistence of immature, unstable spines. This dysregulation disturbs the fine balance of synaptic remodeling, undermining the development of precise and efficient neural circuits. Consequently, inflammation-driven microglial dysfunction during sensitive developmental periods such as adolescence can contribute to cortical thinning and executive function impairments observed in disorders like ADHD.

Importantly, depleting microglia in adult rodents did not produce persistent cognitive impairments, indicating adolescence is a sensitive period during which microglia critically shape executive circuitry (Parkhurst et al. 2022). This study also highlighted microglia's dual role: while excessive activation can be neurotoxic, their normal function is essential for healthy brain maturation. Dysregulated microglial activation, whether excessive or deficient, can impair synaptic connectivity and cortical maturation, offering a biological explanation for how neuroinflammation might lead to structural and functional executive deficits observed clinically (Parkhurst et al. 2022).

A Developmental Pathway to Executive Dysfunction

Taken together, the evidence from imaging, molecular, and mechanistic studies suggests a developmental flow where elevated microglial activation in the frontal lobe could drive abnormal cortical thinning and later executive dysfunction in children. Microglia normally prune synapses, and the loss of synapses manifests as cortical thinning. When this process occurs at the proper developmental pace, cortical thinning strengthens neural networks, increasing efficiency and supporting the maturation of executive functions. However, when microglial activation is excessive or mis-timed, overactive cortical thinning removes necessary synapses and disrupts

dendritic spine maturation, leading to weakened prefrontal circuits. This disrupted pruning pattern produces deficits in memory, planning, attention, and cognitive flexibility. Importantly, while children are more sensitive to changes in microglial activation, a study on frontotemporal dementia (FTD) demonstrated that microglial activation in the elderly predicts executive decline beyond structural atrophy, indicating that inflammation plays a direct role in cognitive deterioration. Pediatric imaging studies reveal that cortical thinning in frontal areas corresponds to impaired attention, working memory, and planning, establishing the structural-functional link. Animal model research provides mechanistic validation that microglial disruption during adolescence causes lasting executive dysfunction by interfering with cortical maturation (Bevan-Jones et al. 2020).

This cascade supports the hypothesis that in children with early cortical thinning, microglial activation may precede and predict cognitive decline, serving as an early biomarker of vulnerability. Elevated microglial activity could either induce excessive synaptic pruning, leading to cortical thinning, or result from an abnormal inflammatory response exacerbating thinning and cognitive deficits. Either way, the interplay between microglial activation and cortical thinning appears integral to the developmental trajectory of executive function (Bevan-Jones et al. 2020; Narr et al. 2009; DiSabato et al. 2016).

Lastly, exploratory analyses integrating immunological and genetic data point to potential modulatory influences on microglial activity. Preliminary associations were observed between pro-inflammatory cytokine levels, specific gene variants related to immune signaling (like *CX3CR1*: a gene that helps control how microglia, the brain's immune cells, move and communicate with neurons, especially during brain development and inflammation), and cortical thinning trajectories (Rogers et al. 2011). These multimodal findings suggest that microglial activation is shaped by a combination of neurodevelopmental timing, immune signaling, and potentially inherited susceptibility. Collectively, these results support the hypothesis that microglial activation in the frontal cortex may serve as an early indicator of disrupted executive circuit development in pediatric populations (Luciana et al. 2012).

Methods

Multimodal Evidence Integration

To explore whether elevated microglial activation in the frontal lobe can predict cognitive decline in children with early cortical thinning, this review collected findings from three key types of studies: adult clinical imaging, pediatric neurodevelopmental imaging, and animal models. These sources were selected to represent both mechanistic insight and real-world relevance.

Literature Search Strategy

The papers were found by a focused search on PubMed and Web of Science using terms like “microglial activation frontal cortex PET,” “cortical thinning ADHD children,” and

“microglia depletion adolescent mice prefrontal cortex.” Papers were included if they directly examined the relationship between brain inflammation, cortical structure, and executive function. Three foundational studies were selected: a PET imaging study on microglia in adults with frontotemporal dementia, several pediatric MRI studies on cortical thinning and cognitive outcomes, and an animal study examining how microglial depletion affects brain development and behavior.

Adult Imaging Data: Microglial Activation as Early Predictor

The first study, by Bevan-Jones et al., looked at adults with frontotemporal dementia (FTD), a condition marked by a significant decline in executive abilities like planning and attention. The researchers used PET imaging to detect active microglia in the brain, specifically in the frontal and temporal lobes (2020). They also used MRI to measure cortical atrophy. Over several years, participants completed regular cognitive assessments. The researchers then analyzed whether early levels of microglial activation could predict how quickly someone’s attention, planning, or memory worsened. Interestingly, microglial activity predicted decline even when brain atrophy was not yet severe, suggesting inflammation could be an early warning sign.

Pediatric Neuroimaging Evidence: Cortical Thinning and Executive Dysfunction

Next, I considered findings from pediatric MRI studies. Narr et al. examined children with ADHD and found that thinning in prefrontal areas of the brain was closely tied to working memory and attention difficulties (2009). McLaughlin et al. focused on children exposed to early adversity and found that those with thinning in frontal brain regions often showed poor executive function skills (2014). Shaw et al. tracked cortical development over time and discovered that delays in prefrontal thinning were linked to symptoms of hyperactivity and impulsivity (2011). Luciana et al. looked at typically developing youth and found that a thicker cortex in the right medial frontal gyrus was associated with better planning skills (2012). All of these studies used structural MRI and common analysis pipelines like FreeSurfer to measure cortical thickness and correlate it with cognitive test scores.

Animal Model Insight: Microglial Function in Brain Development

To better understand how microglia influence development, Parkhurst et al. conducted a rodent study in which they depleted microglia in adolescent mice (2022). Using pharmacological and genetic tools, they removed microglia specifically from the prefrontal cortex during this key developmental window. Later in adulthood, the mice were tested on tasks that measured executive functions like planning and cognitive flexibility. Mice that had lost microglia during adolescence showed long-lasting problems, while mice that lost them as adults did not. The team also examined the physical structure of the brain and found that microglia depletion interfered with the normal pruning of synapses and maturation of dendrites. This supports the idea that microglia are critical for healthy brain development, not just inflammation.

DISCUSSION

Summary of Key Findings

Mechanisms of ADHD and other executive function disorders remain poorly understood, and identifying biological pathways is critical for advancing prevention and treatment. This paper considered evidence from adult neuroimaging, pediatric developmental studies, and rodent models to explore the role of microglial activity in executive dysfunction. Despite differences in methodology across studies, including PET imaging in frontotemporal dementia, MRI-based measures of cortical thickness in children, and experimental microglial depletion in mice, a consistent pattern emerges. Inflammation in the prefrontal cortex, mediated by microglia, appears to shape cortical thinning, which in turn impacts key cognitive outcomes such as planning, attention, and working memory. Together, these findings highlight microglial regulation of cortical maturation as a potential mechanism linking early brain development with later executive dysfunction.

Linking Neuroinflammation, Cortical Development, and Executive Function in ADHD

Elevated microglial activity has been identified as a potential early contributor to neurodevelopmental disorders. In adults with frontotemporal dementia, higher frontal microglial activation predicted subsequent declines in executive function, even before significant cortical thinning occurred (Bevan-Jones et al. 2020). In rodent models, microglial depletion during adolescence, but not during adulthood, led to long-term deficits in planning, working memory, and cortical structure (Parkhurst et al. 2022). Pediatric studies further link cortical development with executive function. Children with ADHD show widespread prefrontal cortical thinning associated with impairments in attention and working memory (Narr et al. 2009). Early adversity predicts similar thinning and subsequent executive dysfunction (McLaughlin et al. 2014). Delayed cortical maturation correlates with greater hyperactivity and impulsivity (Shaw et al. 2011), while cortical thickness in the right medial frontal gyrus predicts planning ability in typically developing children (Luciana et al. 2012).

The novelty of synthesizing these findings is the integration of microglial biology, neuroimaging markers of cortical thinning, and behavioral outcomes related to ADHD symptoms, such as attention, memory, and planning deficits. This integrated framework proposes that early microglial activation may drive abnormal synaptic pruning and cortical thinning in prefrontal regions, which in turn predicts executive function impairments characteristic of ADHD in adolescence. By bridging longitudinal human imaging with mechanistic animal studies, this perspective advances our understanding of how neuroimmune processes contribute to ADHD's neurodevelopmental trajectory. This new approach highlights the potential of targeting microglial activity and neuroinflammation as early biomarkers and therapeutic points of intervention to alter the course of executive dysfunction in at-risk children.

Current Gaps and Methodological Challenges

Still, there are key gaps to fill. Direct studies in children that measure both microglial activation and cortical thinning are currently lacking, largely due to the difficulty of conducting PET scans in pediatric populations. Advances in non-invasive imaging and safer tracer compounds will be crucial to moving forward. Another challenge is the complexity of microglial behavior. Microglia support brain development in many ways, and it is not yet clear how to distinguish between healthy and harmful activation in living humans. More specific biomarkers are needed to tell when inflammation becomes a threat to development.

Environmental and genetic factors also deserve more attention. Research by Katie A. McLaughlin and colleagues suggest that early adversity can shape brain development, possibly through effects on immune signaling (2014). However, it is still unclear how those life experiences interact with individual differences in inflammation or microglial function. Future work should explore these interactions by combining neuroimaging with genetic and immunological testing.

Implications for Early Detection and Intervention

Even with these limitations, this research has meaningful implications. If inflammation in the frontal cortex can act as an early signal of executive dysfunction, it could lead to new strategies for identifying children at risk. That, in turn, could allow for earlier and more targeted interventions to support healthy cognitive development. Specifically, targeting microglia, the brain's resident immune cells, may be a promising approach, as these cells play a critical role in shaping synaptic development during sensitive periods. Microglia actively monitor the brain environment and regulate synaptic pruning, a process essential for refining neural circuits and ensuring optimal connectivity, particularly in the prefrontal cortex. Inflammatory signals can alter microglial function, causing them to become either overactive or underactive. Such dysregulation may result in abnormal synaptic pruning, leading to either excessive elimination of synapses or insufficient removal of weak or unnecessary connections. Both scenarios can disrupt the formation of efficient neural networks that underlie executive functions like planning, working memory, and cognitive flexibility. By understanding how inflammation influences microglial activity and synaptic remodeling, researchers and clinicians can develop interventions aimed at modulating neuroinflammation or restoring microglial homeostasis. This could help preserve healthy synaptic development and protect against long-term cognitive impairments, particularly if applied during key developmental windows such as adolescence. Overall, these findings underscore the potential of microglia as a therapeutic target to mitigate the impact of early-life environmental stressors on brain development and executive function.

In summary, the combined findings from adult imaging, pediatric neurodevelopment, and rodent models suggest that elevated microglial activation in the frontal cortex is a strong candidate mechanism for predicting cognitive decline in children with early cortical thinning. Continued research in this area may lead to more effective tools for early detection and interventions that preserve executive function during critical developmental periods.

Works Cited

- Bevan-Jones, William R., et al. "Neuroinflammation and Protein Aggregation Co-Localize across the Frontotemporal Dementia Spectrum." *Brain*, vol. 143, no. 3, 2020, pp.1010–1026. Oxford University Press, <https://doi.org/10.1093/brain/awaa033>.
- Dickerson, Bradford C., et al. "The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals." *Cerebral Cortex*, vol. 19, no. 3, 2009, pp. 497–510, Oxford University Press, <https://doi.org/10.1093/cercor/bhn113>.
- DiSabato, Damon J., et al. "Neuroinflammation: The Devil Is in the Details." *Journal of Neurochemistry*, vol.139, no.S2, 2016, pp.136–153, . <https://doi.org/10.1111/jnc.13607>.
- Luciana, Monica, et al. "Cortical Thickness in the Right Medial Frontal Gyrus Predicts Planning Performance in Healthy Children and Adolescents." *Frontiers in Human Neuroscience*, vol.6, 2012, article101, <https://doi.org/10.3389/fpsyg.2023.1196707>.
- McLaughlin, Katie A., et al. "Widespread Reductions in Cortical Thickness Following Severe Early-Life Deprivation: A Neurodevelopmental Pathway to ADHD." *Biological Psychiatry*, vol. 76, no. 8, 2014, pp. 629–638, <https://doi.org/10.1016/j.biopsych.2013.08.016>.
- Narr, Katherine L., et al. "Widespread Cortical Thinning Is a Robust Anatomical Marker for Attention-Deficit/Hyperactivity Disorder." *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 48, no. 10, 2009, pp. 1014–1022, <https://doi.org/10.1097/CHI.0b013e3181b395c0>.
- Schalbetter, S. et al. "Adolescence Is a Sensitive Period for Prefrontal Microglia to Act on Cognitive Development." *Science Advances*, vol. 8, no. 10, 2022, eabi6672, <https://doi.org/10.1126/sciadv.abi6672>.
- Rogers, J. T., et al. "CX3CR1 Deficiency Leads to Impairment of Hippocampal Cognitive Function and Synaptic Plasticity." *Journal of Neuroscience*, vol. 31, no. 45, 2011, pp. 16241–16250, <https://doi.org/10.1523/jneurosci.3667-11.2011>.
- Shaw, Philip, et al. "Cortical Development in Typically Developing Children with Symptoms of Hyperactivity and Impulsivity: Support for a Dimensional View of ADHD." *Proceedings of the National Academy of Sciences*, vol. 108, no. 49, 2011, pp. 10358–10363, <https://doi.org/10.1176/appi.ajp.2010.10030385>.

Fast Radio Bursts: An Introduction to One of Astronomy's Most Mysterious Signals

By Rachelle Pinkerton

Abstract

Fast Radio Bursts (FRBs) are incredibly powerful, millisecond-long bursts of radio waves that emit as much energy in a fraction of a second as the Sun does in an entire day. Since their discovery by Lorimer and Narkevic in 2007, FRBs have challenged astronomers' understanding of neutron stars, magnetars, and the environments in which these extreme objects form. This paper introduces FRBs, how they are detected, and the two main types—repeating and non-repeating bursts. It also traces the history of FRB research from the Lorimer Burst, through the “perlyton” confusion era, to the discovery of the first repeating FRB, FRB 121102. A major focus is placed on recent discoveries from the CHIME telescope, whose wide field of view and high detection rate have revolutionized the field by uncovering hundreds of FRBs, identifying new repeating behavior, and confirming the first FRB-like burst from a magnetar in the Milky Way. Finally, this paper discusses how FRBs are used today to study intergalactic matter, magnetic fields, star formation, and fundamental physics. Even though their origins remain uncertain, FRBs have already become essential tools in modern astrophysics.

Introduction

Fast Radio Bursts (FRBs) are broadband, incredibly powerful bursts of radio waves that last for only a few milliseconds. Duncan Lorimer, a professor of physics and astronomy at West Virginia University, and his student David Narkevic discovered the signals in 2007 while analyzing data from the Parkes radio telescope in Australia. FRBs emit as much energy in a fraction of a second as the sun does in an entire day (Tinelmis, 2025). FRBs have dispersion, and their brightness lies within the 400-8000 MHz frequency range. FRBs are estimated to occur approximately 1000 times per day (Singh et al, 2025).

FRBs were previously thought to originate from young neutron stars that expend a lot of energy as they're forming. Recently, with the CHIME outtrigger project, new discoveries have shown that these signals aren't just coming from neutron stars. One signal, according to Shah et al (2024), was found in a quiescent galaxy: FRB 20240209A. Additionally, due to the brief nature and distance of FRBs, astronomers could only approximate an originating area. With FRB 20250316A, astronomers were able to pinpoint the galaxy from which it originated.

FRBs are important to astronomy in many ways. From what Lorimer et al (2024) tell us, fast radio bursts are important for two main reasons. First, FRBs have forced astronomers to rethink what neutron stars and magnetars can do, revealing objects that are brighter and more extreme than anything known before. FRBs are found in different environments such as dwarf galaxies, spiral arms, and even globular clusters. They also push work on things like ultra-long-period magnetars, white dwarf mergers, and other exotic systems. Second, even though their sources aren't fully understood, they are being used to study the gas between

galaxies and the distribution of normal matter in the universe. FRBs act like “intergalactic dipsticks” to measure the electrons along different sight-lines through the universe.

What are Fast Radio Bursts

FRBs are very fast, very bright bursts of radio waves that are thought to come from extremely compact objects such as neutron stars and possibly even black holes. These bursts carry an enormous amount of energy and range from a fraction of a millisecond to 3 seconds. They originate from beyond our own galaxy, often billions of light-years away (Reischke and Hagstotz, 2022).

FRBs are detected by processing radio telescope data through a series of steps. The telescope collects radio waves and focuses them onto receivers. The receivers then measure signal strength at many different frequencies and at very fast time intervals. These measurements form a data cube: time, frequency, and intensity. The human-made radio signals from Wi-Fi, satellites, and phones, to name a few, must then be cleaned out as they can also look like astrophysical bursts. Since the signals arrive at different times with high frequencies first and low frequencies later, the astronomers must then apply many different dedispersion measures to find which one lines up the signals. The dispersed data is then summed into time series, smoothed, normalized, and searched using matched filters to identify peaks above the noise. The signal is then run through further checks to reject false positives and the remaining candidates are classified as true FRB detections. (Petroff et al, 2019, 4)

There are two main types of FRBs: Repeating and non-repeating. Repeating FRBs produce multiple bursts over time and tend to have more complex pulse structures, narrower bandwidths, and longer durations. Non-repeating, or one-off FRBs, are often brighter, have an extremely short duration, lack the structured appearance of repeating bursts, and are possibly due to catastrophic events. (Petroff et al, 2019; CHIME/FRB Collaboration, 2019; Fonseca et al, 2020; Pleunis et al, 2021)

The Discovery and History of FRBs

In 2007, while looking through archival pulsar survey data, Duncan Lorimer and his student David Narkevic identified a bright flash of radio waves, lasting less than 5 milliseconds in duration. The burst originated in 2001 and its brightness was registered at $\sim 800 \pm 400$ Jy ms, (Lorimer et al, 2007). Initially, this was thought to be a unique event, but as more FRBs were detected, it became apparent that the Lorimer Burst was not the only one of its kind. This discovery proved to be a significant milestone in the field of radio astronomy as it provided astronomers with a new class of objects and sparked an interest in the identification of their sources and the development of new theories about their origins and properties.

In 2010, as part of her PhD, Sarah Burke Spolaor set out to try to find more FRBs by using the same technique as Lorimer. Instead, she found strange new signals that she named “peryttons”. The initial peryttons found were similar to the Lorimer burst in that the signal arrived at different times at different wavelengths, but the span of it covered the entire receiver instead of

just a single fixed point. This showed that the signals were not coming from space, they were coming from Earth. An RFI monitor was installed at the Parkes site that would aid in the discovery of what these mysterious signals were. It revealed that at the time of each new peryton event, there was a period of radio emission at the frequency 2.5 GHz. Since many consumer electronics emit at that frequency, the team set about testing the microwave ovens at the site. Almost two months after the tests began, they found they could generate perytons by stopping the ovens by opening the door and having a direct line of sight between the microwave and the telescope receiver (Petroff, 2015).

FRB 121102 was first detected on November 2, 2012. This was the biggest turning point in FRB history as it proved that not all FRBs came from catastrophic events, it split FRBs into two categories: repeating and non-repeating, and allowed astronomers the ability to pinpoint an origin location as a dwarf galaxy more than three billion light-years from Earth. (Chatterjee, 2017), (Spitler et al, 2014). Follow-up observations found ten additional bursts from the same direction which proved that its source survived the events that caused the bursts. In total, 41 bursts were detected from 2017-2020, and may have come from two separate sources (Cruces et al, 2020).

CHIME and Recent Discoveries

The Canadian Hydrogen Intensity Mapping Experiment (CHIME) is a radio telescope with no moving parts. It was originally built to map hydrogen and is optimized to have a large field of view (~200 square degrees) and broad frequency coverage (400-800 MHz). It has a unique, custom, cylindrical telescope design, which allows for its enormous field of view and good sensitivity in both linear polarizations.

CHIME completed construction of three new radio telescopes to work alongside the main CHIME instrument in 2022 (Gilliard, 2022). The additions to the array have allowed astronomers to detect 10-100 times more FRBs with better time resolution.

CHIME's Significant Contributions

CHIME revolutionized FRB astronomy by the number of FRB's they detected every year - hundreds to thousands of fast radio bursts. Their first catalogue in 2021, before the outriggers were constructed, contains 535 FRBs. Their later catalogs contain even more showing that FRB's aren't rare exotic events. CHIME now produces the majority of the world's FRB detections. Before CHIME, only two repeaters were known to exist. CHIME increased that number dramatically. A couple of major findings include FRB 180916, FRB 190520, and FRB 20201124A.

In 2018, CHIME discovered a repeating FRB with a twist. Most repeating FRBs burst unpredictably, but FRB 180916.J0158+65 showed something brand new. It repeats on a precise 16.35-day cycle. It is active for about 4-5 days and then is followed by ~11-12 days of silence before it repeats (Chawla et al, 2020). This was the first-ever periodic activity window discovered in an FRB.

A repeating 16-day cycle cannot be caused by a random process which strongly suggests this could be caused by a binary system, a precessing magnetar, or a source interacting with stellar winds or orbital environments.

On April 28, 2020, an FRB was detected from the known magnetar SGR 1935+2154. This FRB, later designated FRB 200428, was an extremely bright, millisecond-duration radio burst between roughly 400 MHz and 800 MHz. This was the first FRB-like burst detected within the Milky Way and linked to a known astrophysical object. The detection confirmed that radio bursts can originate from magnetars, strengthening magnetar models for FRB sources (Andersen et al, 2020).

How FRBs Are Used Today

Petroff et al (2022) tells us that even though FRBs origins remain uncertain, fast radio bursts have become powerful tools for studying the universe. Astronomers now use FRBs to map the distribution of matter between galaxies, measure cosmic magnetic fields, test fundamental physics, and probe the environments around neutron stars and magnetars. Their extreme brightness and rapid duration make them excellent “cosmic flashlights” for studying the gas and plasma that lie along their path. As telescopes like CHIME detect thousands of new bursts, FRBs are becoming essential for cosmology, galaxy evolution research, and real-time multi-wavelength astronomy. Even without fully understanding what produces them, FRBs are already shaping modern astrophysics.

Most matter is spread out as thin plasma between galaxies. As an FRB travels through space, its radio waves get dispersed. The amount of dispersion tells astronomers how much matter the burst traveled through. By comparing FRBs at different distances, we can map the distribution of this hidden, intergalactic material. They are the best tool we have for measuring gas between galaxies.

FRBs can also be used to study host galaxies and star formation with the ability to now localize FRBs to specific galaxies. Many repeaters are found in star-forming regions, helping scientists study how young stars evolve into magnetars, and allow astronomers to probe magnetic environments inside galaxies.

Since FRBs can travel over billions of light-years, astronomers can use FRBs to test whether all types of light travel at the same speed (a test of Einstein’s Weak Equivalence Principle) and whether the laws of physics change over time or distance. They can test plasma physics under extreme magnetic fields and also how radiation moves through highly magnetized environments.

FRBs often have a strong polarization, meaning the radio waves are aligned in certain directions. This lets astronomers measure magnetic fields in galaxies, track how magnetic fields change near neutron stars, and study turbulence in the intergalactic medium.

Many FRBs likely come from magnetars. Each burst teaches us more about how magnetars release energy and how their magnetic fields evolve, and can also teach us how massive stars die and collapse into neutron stars.

FRBs are extremely bright and short, like cosmic camera flashes. When they pass through gas clouds, galaxy halos, plasma streams, or stellar winds, their signals get altered. Astronomers can analyse these changes to study the stuff between us and the burst.

Conclusion

Fast Radio Bursts remain one of astronomy's most compelling mysteries. Even with all the progress made within the last two decades, they continue to defy easy explanation. What began as a single signal in 2007 found in archival data from 2001 has grown into a field full of surprises, contradictions, and breakthroughs. FRBs are brief, bright, and in most cases impossibly distant, yet the amount of information they provide is reshaping how we think about neutron stars, magnetars, and the matter hidden between galaxies.

From the Lormier Burst to the "Peryton" era, from the first repeater to CHIME's flood of detections, the history of FRBs has been increasingly abundant in its teachings. When astronomers thought they had an idea of what an FRB is, a new observation would require reassessment. The discovery of repeaters forced them to rethink the original hypothesis that all FRBs came from catastrophic events. The identification of periodic activity cycles forced astronomers to need to consider new physical scenarios. The Milky Way magnetar burst settled the doubt that magnetars could produce FRB-like emissions while also simultaneously opened the door to questions we still can't answer.

CHIME's contributions have transformed the field entirely. With its large field of view and high detection rate, more and more FRBs are being detected and cataloged. They have proven that FRBs are not rare cosmic events, but that they are frequent, powerful, and extremely useful. Their ability to travel across billions of light-years makes them perfect cosmic probes to expose the universe's secrets.

Even now, we still don't fully understand them, what produces them, why some repeat while others do not, or what controls the timing of periodic bursts. But therein lies the excitement. Each burst is a reminder that the cosmos is far vaster, stranger, and more intricate than we had previously assumed.

FRBs are powerful, informative, and currently unpredictable. In the years ahead, with telescopes like CHIME pushing the limits, FRB astronomy is poised to become one of the central tools in unraveling the universe's structure. The universe has plenty of surprises left and we should keep our eyes open.

Works Cited

- Andersen, B., et al. A Bright Millisecond-Duration Radio Burst from a Galactic Magnetar. 2020. Accessed 14 Dec. 2025.
- Burke-Spolaor, S., et al. "RADIO BURSTS with EXTRAGALACTIC SPECTRAL CHARACTERISTICS SHOW TERRESTRIAL ORIGINS." *The Astrophysical Journal*, vol. 727, no. 1, Dec. 2010, p. 18, <https://doi.org/10.1088/0004-637x/727/1/18>. Accessed 12 Sept. 2020.
- Chatterjee, Shami. "Focus on the Repeating Fast Radio Burst FRB 121102 - the Astrophysical Journal Letters - IOPscience." *Iop.org*, 2017, iopscience.iop.org/journal/2041-8205/page/Focus_on_FRB_121102. Accessed 14 Dec. 2025.
- Chawla, P., et al. "Detection of Repeating FRB 180916.J0158+65 down to Frequencies of 300 MHz." *The Astrophysical Journal*, vol. 896, no. 2, IOP Publishing, June 2020, pp. L41–41, <https://doi.org/10.3847/2041-8213/ab96bf>. Accessed 13 Aug. 2023.
- Chen, Ge, et al. "A Comprehensive Observational Study of the FRB 121102 Persistent Radio Source." *The Astrophysical Journal*, vol. 958, no. 2, IOP Publishing, Nov. 2023, pp. 185–85, <https://doi.org/10.3847/1538-4357/ad02f3>. Accessed 13 Dec. 2025.
- Cruces, M., et al. "Repeating Behaviour of FRB 121102: Periodicity, Waiting Times, and Energy Distribution." *Monthly Notices of the Royal Astronomical Society*, vol. 500, no. 1, Oct. 2020, pp. 448–63, <https://doi.org/10.1093/mnras/staa3223>. Accessed 11 July 2023.
- Fonseca, Eduardo, et al. "Nine New Repeating Fast Radio Burst Sources from CHIME/FRB." *The Astrophysical Journal*, vol. 891, no. 1, IOP Publishing, Feb. 2020, pp. L6–0, <https://doi.org/10.3847/2041-8213/ab7208>. Accessed 28 Apr. 2023.
- FRB Collaboration, et al. "FRB 20250316A: A Brilliant and Nearby One-off Fast Radio Burst Localized to 13 Parsec Precision." *ArXiv.org*, 2025, arxiv.org/abs/2506.19006. Accessed 13 Dec. 2025.
- Gilliard, Geoff. "CHIME Outrigger Telescopes Boost Search for Fast Radio Bursts." *UBC Science*, 2022, science.ubc.ca/news/chime-outrigger-telescopes-boost-search-fast-radio-bursts. Accessed 14 Dec. 2025.
- Hensley, Kerry. "A Unified Explanation for Fast Radio Bursts." *AAS Nova*, 10 Dec. 2025, aasnova.org/2025/12/10/a-unified-explanation-for-fast-radio-bursts/. Accessed 13 Dec. 2025.
- Lorimer, D. R., et al. "A Bright Millisecond Radio Burst of Extragalactic Origin." *Science*, vol. 318, no. 5851, Nov. 2007, pp. 777–80, <https://doi.org/10.1126/science.1147532>. Accessed 12 Sept. 2020.
- Lorimer, Duncan R., et al. "The Discovery and Significance of Fast Radio Bursts." *ArXiv (Cornell University)*, vol. 369, no. 59, Cornell University, May 2024, <https://doi.org/10.1007/s10509-024-04322-6>.

- Petroff, E., et al. “Fast Radio Bursts.” *The Astronomy and Astrophysics Review*, vol. 27, no. 1, May 2019, <https://doi.org/10.1007/s00159-019-0116-6>.
- . “Identifying the Source of Perytons at the Parkes Radio Telescope.” *Monthly Notices of the Royal Astronomical Society*, vol. 451, no. 4, June 2015, pp. 3933–40, <https://doi.org/10.1093/mnras/stv1242>. Accessed 12 Sept. 2020.
- Petroff, Emily. “How We Found the Source of the Mystery Signals at the Dish.” Swinburne.edu.au, 25 May 2015, www.swinburne.edu.au/news/2015/05/how-we-found-the-source-of-the-mystery-signals-at-the-dish/. Accessed 13 Dec. 2025.
- Pleunis, Ziggy, et al. “Fast Radio Burst Morphology in the First CHIME/FRB Catalog.” *The Astrophysical Journal*, vol. 923, no. 1, IOP Publishing, June 2021, pp. 1–0, <https://doi.org/10.3847/1538-4357/ac33ac>. Accessed 28 Apr. 2023.
- Reischke, Robert, and Steffen Hagstotz. “Consistent Constraints on the Equivalence Principle from Localized Fast Radio Bursts.” *Monthly Notices of the Royal Astronomical Society*, vol. 523, no. 4, Oxford University Press (OUP), June 2023, pp. 6264–71, <https://doi.org/10.1093/mnras/stad1866>. Accessed 13 Dec. 2025.
- Shah, V., et al. “A Repeating Fast Radio Burst Source in the Outskirts of a Quiescent Galaxy.” ArXiv.org, 2024, arxiv.org/abs/2410.23374. Accessed 13 Dec. 2025.
- Sharma, Ritesh, et al. “Fast Radio Bursts through Radio Astronomy.” *The European Physical Journal Special Topics*, vol. 234, no. 16, Springer Science+Business Media, Mar. 2025, pp. 5093–104, <https://doi.org/10.1140/epjs/s11734-025-01571-6>. Accessed 13 Dec. 2025.
- Spitler, L. G., et al. “FAST RADIO BURST DISCOVERED in the ARECIBO PULSAR ALFA SURVEY.” *The Astrophysical Journal*, vol. 790, no. 2, July 2014, p. 101, <https://doi.org/10.1088/0004-637x/790/2/101>. Accessed 12 Sept. 2020.
- The CHIME/FRB Collaboration. “A Second Source of Repeating Fast Radio Bursts.” *Nature*, vol. 566, no. 7743, Springer Science and Business Media LLC, Jan. 2019, pp. 235–38, <https://doi.org/10.1038/s41586-018-0864-x>. Accessed 13 Dec. 2025.
- Thornton, D., et al. “A Population of Fast Radio Bursts at Cosmological Distances.” *Science*, vol. 341, no. 6141, July 2013, pp. 53–56, <https://doi.org/10.1126/science.1236789>. Accessed 12 Sept. 2020.
- tinelmis. “Fast Radio Burst - Definition & Detailed Explanation - Astronomical Objects Glossary - Sentinel Mission.” Sentinel Mission, 12 Oct. 2025, sentinelmission.org/astronomical-objects-glossary/fast-radio-burst/#google_vignette. Accessed 13 Dec. 2025.

Battling Institutional Racism: The Importance of Activist Messaging in Combatting Social Separation

By Mengchen Zhang

Abstract

As immigrants continue to come to the US, as minority groups grow larger, the US continues to grapple with its history of institutional racism. In recent times, there has been a surge in visibility for certain movements like the Black Lives Matter movement and Stop AAPI Hate, in light of police brutality. With the current increase in visibility, now is the perfect time for activist organizations to collaborate and enact meaningful change. However, societal tactics like outdated ideology and the concept of anti-Blackness forbid two of the largest minorities from collaborating: the AAPI community and the Black community. When the two communities do interact and support each other, these actions are often overlooked by history. This study will examine the effects of separation tactics, as well as possible solutions, both past and present, to addressing this issue.

Introduction

In recent times, groups have become increasingly polarized across a wide variety of issues, creating a divided environment. In terms of activism, this is detrimental to the success of a movement. Activism relies on the support from a significant portion of the public to force a government to enact change. When it comes to battling institutional racism, solidarity among activists and organizations can be an act of protest in itself against the conflict society imposes between minorities. The biggest example of this is the divide between the Asian American community and the African American community. Society utilizes multiple tactics in order to keep the two largest minority groups apart, driving wedges between them and keeping both groups from enacting meaningful change.

Tactics of Social Separation

There are many strategies ingrained in our current social structure to separate the AAPI(Asian American Pacific Islander) community from the Black community, creating tension. This tension further complicates these groups' abilities to find common ground, despite being one of the largest minority populations in the United States.

Race: Theory or Fact?

Race is often treated as something steeped in biology and science, a clear marker of who "belongs" and who doesn't. The truth is, race is not a concrete factor—it's actually a composite term that can be separated into three categories: color, ancestry, and class.

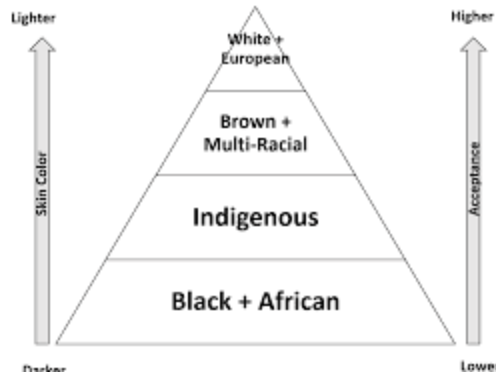


Fig 1. Example of racial hierarchy

However, when we treat race as absolute and singular, it divides minority groups and their subsequent causes. Although many minorities share a history of being oppressed in US society, treating race as an unchangeable component allows for the fragmentation of social movements. Each cause becomes highly individualized and activists focus on their separate interests rather than trying to find common struggles across the differences.

Being able to find points of connection and/or similarity across contrasting social circumstances is essential to building alliances and solidarity. Without that ability, activists have no way to demonstrate widespread support across diverse groups of people, which in turn affects the “success” of their causes. As politicians and people of power tend to only listen when they can no longer ignore the public’s voice, it becomes essential for activists to show support for their cause from people across the board.

“Good” vs. “Bad” Stereotypes

Everyone has stereotyped someone or been stereotyped before; it is a given in a society where implicit bias exists. However, stereotypes affect minority groups in different ways and can further exacerbate pre-existing tensions between groups.

While stereotyping is inherently harmful for the way it generalizes a group of people, different stereotypes can be perceived as “advantageous” or “disadvantageous” for the groups they apply to. Such perceptions can affect relationships between minority groups, overlooking the fact that these conflicts are often created explicitly to prevent two or more minorities from collaborating.

There are many stereotypes applied to Asian Americans, like the “forever foreigner,” but the most commonly perpetrated stereotype is the “Model Minority Myth.” The “Model Minority Myth” can be seen as a “favorable” stereotype because it assumes that all Asians are high achievers and well-off financially. One of the most common “versions” of the stereotype is the belief that all Asians are good at STEM-related subjects.

Asian adults' experiences with 'model minority' stereotypes vary across education and income

% of Asian adults who say each of the following has happened in their day-to-day encounters with strangers in the U.S.

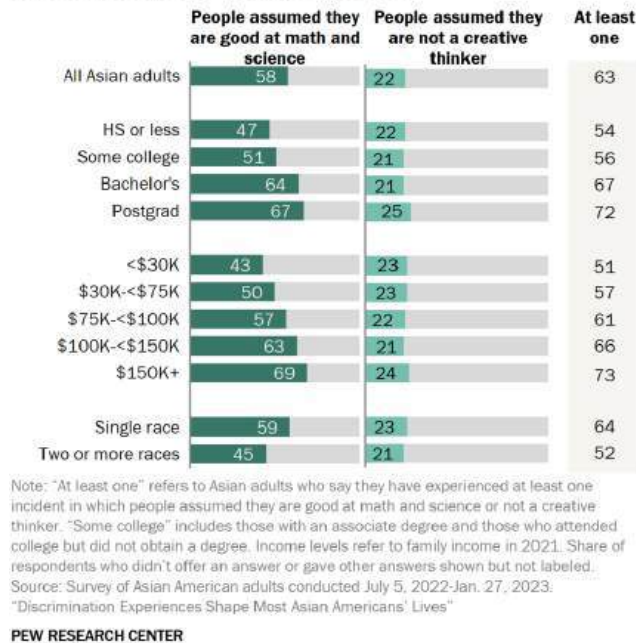


Fig 2. Data from Pew Research Center

Alongside the harm that comes from this generalization of the AAPI community, which includes ignoring and overlooking those who don't fit the false belief, the stereotype also drives a wedge between the Asian American community and the African American because it indirectly compares Asian Americans to Black people. By insinuating that Asian Americans are the "good" minority, Black people become the "bad" minority – in this case, lazy and poor. This distinction can lead to misunderstanding and resentment between the two groups, as the stereotype can be (and has been) used to promote one minority group's cause while simultaneously used to deny another group's cause.

Racial Triangulation and the Concept of Anti-Blackness

America's perception of race has often been split into Black and White, with no room for other races to exist. Racial triangulation addresses how Asian Americans fit into this supposedly clear split between racial groups.

According to Claire Jean Kim, Asian Americans throughout history have been assigned closeness to Whites when it was advantageous for politicians to do so—like when Asian Americans were afforded civil rights during the Vietnam War. Asian refugees were flocking to the US while it was simultaneously trying to form alliances with governments in Asia; in a way, this forced the US to change its racial politics around Asian immigrants and Asian Americans.

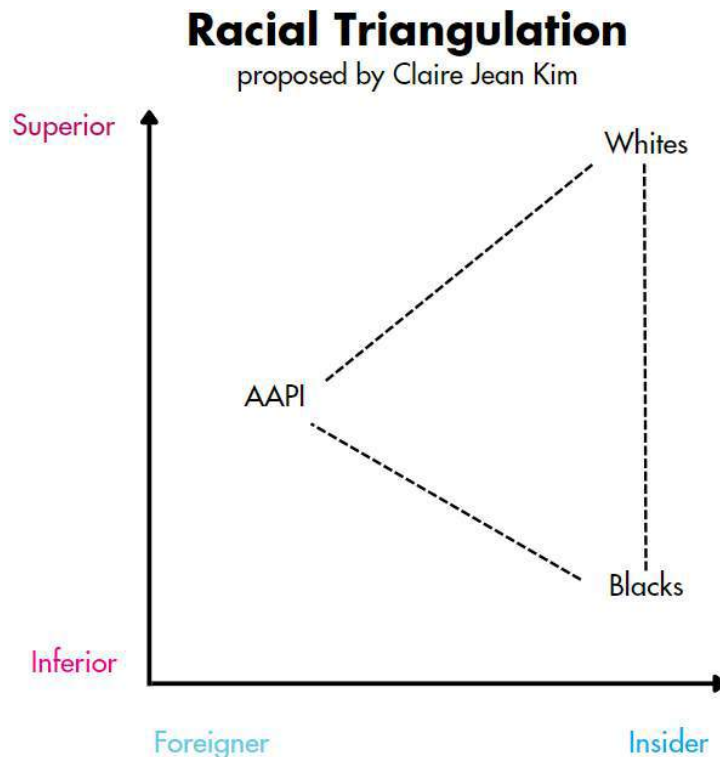


Fig 3. Visualization of racial triangulation

However, Asian Americans have also been lumped with Blacks – again, when it was convenient for politicians to do so. This tactic has also been used to deny both groups their rights. By presenting Asian Americans as a political weapon, using the community to pick and choose whose rights are more important, it continues to drive wedges between the Asian and Black communities. By pitting these two groups against each other, the dominant group, white people, can remain in power and reinforce white supremacy.

This concept was first presented in 1999; since then, Claire Jean Kim, the author, has updated her research to examine the concept of anti-Blackness rather than racial triangulation. Both theories are extremely similar because racial triangulation can be categorized under the broader umbrella of anti-Blackness. The concept of anti-Blackness essentially states that while white is superior, the most important thing is to be not-Black. By being not-Black, you are automatically superior to Black people while simultaneously inferior to white people. This is one of the main mechanisms that has fueled US racial hierarchies, as well as many in the Western world. The significance of anti-Blackness lies in the fact that it can operate at both individual and systemic levels, from everyday interactions to institutional policies.

Role of Messaging in Activism

Although messaging in activism is important for raising awareness for individual movements, it's also essential towards building alliances. By showing visual and/or verbal support for another activist group and their cause, both groups can work towards achieving their goals together.

Messaging in activism is primarily used in relation to memory. Memory is vital to the effectiveness of messaging, as well as garnering political action and resistance. Activist communication can utilize memory to spur people to take action. For example, pro-democracy activists in Hong Kong and the Chinese diaspora use the 1989 Tiananmen Square massacre to fight against the forced forgetting of the Chinese government. Using the memory of a significant event to encourage people to take action is known as “memory activism.”

Activist communication is also important for creating a collective memory. Collective memory refers to the shared narratives and beliefs of a group regarding their past; this then affects their view of the present and the future. State-sanctioned memories tend to perpetuate long-held beliefs and systems like institutional racism.

Activists can challenge these state-sanctioned memories by redefining the past or drawing attention to overlooked parts of history. For example, Grace Lee Boggs is frequently portrayed as a key figure in Asian American civil rights, but her contributions to the African American movement are often overlooked. This simplified portrayal of her can lead to misunderstandings between the two minority groups. By cross-posting online about her contributions in both the Black civil rights movement and the fight for Asian American civil rights, we can find more common points of solidarity.

Solidarity

Solidarity, at its simplest definition, is the act of showing up and standing up for someone—or group—whose issues don’t appear to align to your own. In activism, solidarity goes beyond simply showing up; solidarity in activism requires recognizing that another group’s issues are linked to your own. Solidarity in activism requires a shared understanding of historical and modern inequalities.

Activists

This is far from a completed list of activist collaborations; however, the people listed here are well-known figures in the civil rights movement with relationships that are often overlooked, despite their importance. This is also not an extensive list of their work, so if you find an activist intriguing, you may want to look into their work further.

Yuri Kochiyama(1921-2014) was a Japanese-American activist who advocated for multiple causes, including ending job discrimination, the creation of ethnic studies, and supporting anti-war movements. She met Malcolm X(1925-1965) in 1963 – he encouraged her to embrace an internationalist view of civil rights, connecting the struggles of Black people in the US to those of Asian or African descent around the world. Although their collaboration was cut short by Malcolm X’s assassination in ‘65, Kochiyama remained an avid supporter of Black nationalism, the concept of creating a separate nation for Black people, as well as reparations for the internment of Japanese-Americans.



Fig 4. Malcolm X(left) and Yuri Kochiyama(right)

Grace Lee Boggs(1915-2015) was a Chinese-American activist who worked with her husband to advocate for labor, civil rights, and Black power struggles. Her work in activism began with fighting for better housing conditions for her and the Black community living in the same projects in Chicago. She also helped organize important events like the March for Freedom in Detroit, where MLK Jr. gave a version of his “I Have A Dream” speech, and the Grassroots Leadership Conference, where Malcolm X gave his “Message to the Grassroots” speech.



Fig 5. Grace Lee Boggs

Martin Luther King Jr.(1929-1968) is famous for being a key figure during the Civil Rights movement of the 50s and 60s. During this time, Dr. King also occasionally spoke out against the conflict in Vietnam; however, his commitment to the cause did not take off until he received a letter from Thich Nhat Hanh(1926-2022; also known as Thay). Thich Nhat Hanh was a Vietnamese Buddhist monk who urged Dr. King to recognize the parallels between the monks’ nonviolent struggle and the Black fight for human rights, leading to a close relationship. Dr. King also later made several speeches condemning the US’ involvement in Vietnam.



Fig 6. Martin Luther King Jr.(left) and Thich Nhat Hanh(right)

Interracial Activist Groups

The organizations listed below fight for intersectional justice, rather than individualizing their community's struggles. Again, this is far from an extensive list—there are many more organizations working to create a more united world.

Freedom, Inc™ is an activist organization based in Dane County, Wisconsin. It is a Black and Southeast Asian non-profit organization that works with low-income families to challenge the root causes of violence, poverty, and discrimination.

Race Forward™ is a collaborative group that works with other organizations and communities to achieve its goal of a just, multiracial society. Their award-winning news site, *Colorlines*, focuses on pushing the conversation on race in the media.

Showing Up for Racial Justice(SURJ)™ is a national organization that aims to involve white people in the fight for racial and economic justice. The organization works in the deep South and rural communities – places targeted by the far-right – in an effort to break the right's political power.

The Third World Liberation Front™ was a coalition formed from multiple racial and ethnic groups, including African Americans, Asian Americans, and Latinx Americans. They called for the creation of a college specifically dedicated to the histories of marginalized communities. Their work led to the creation of the Department of Ethnic Studies.

Coalition-Building Techniques

Coalition-building is an ongoing process, as groups continually challenge systems of power. It cannot be a “one-and-done” event if we truly want to build alliances between both minority groups. Building alliances requires a continuous process of collaboration and mutual understanding.

Utilizing Social Media

With new technology, activist messaging can reach new levels of outreach. During the civil rights movement in the 60s, activists relied on word-of-mouth and physical advertisements to garner participants in events like marches and fundraisers; today, organizations can make a post or hashtag online to generate attention and support.

Whilst raising awareness for individual causes is important, social media can also be utilized to show support for similar causes. Posts can draw on historical narratives and link them to the present through captions and visuals. They can also be used to create bridges that might be harder to build in real life, like across intergenerational and interracial gaps.

For example, although Yuri Kochiyama and Malcolm X shared a close friendship and supported each other's work, their relationship was never caught on camera. The only photo depicting their bond was taken right after Malcolm X was shot; Kochiyama is depicted kneeling over Malcolm X's body. The post below reimagines the bond between the formidable pair and its impact, showcasing a long history of Black-Asian solidarity.



Fig 7. Instagram post from 18 Million Rising™

Social media can also be used to modernize and reuse old slogans. In the Instagram post below by the Asian American Feminist Collective, the organization repurposes an anti-Asian sentiment to show support for the Black freedom movement, creating the slogan “Yellow Peril Supports Black Power.” Not only does this further highlight solidarity, it also serves as a reclaiming of a slur, empowering the AAPI community.



Fig 8. Instagram post from Asian American Feminist Collective™

Redefining Race

By redefining race, we can challenge and disrupt regimes of power. This new understanding can affect the concept of relational racialization (where one race is seen as superior to another based on racial hierarchies). If we acknowledge race as a theory rather than a fact, we can bridge the gaps between minorities and allow for more collaboration and alliances. Groups can also recognize the similarities in their respective causes, leading to less specialized movements. Although this is a difficult thing to accomplish because race has become so ingrained in our society, there have been attempts, including a workshop that encourages people to reimagine their place in society if race were not a factor. There have also been facilitated discussions online between members of the Asian and Black communities, where they can find similarities in how they think, rather than how society perceives these two minorities.

Works Cited

- "About Us." The Third World Liberation Front, twlf.berkeley.edu/about-us.
- Fu, May, et al. "#Asians4BlackLives: Notes from the Ground." *Amerasia Journal*, vol. 45, no. 2, 2019, pp. 253-270.
- "Grace Lee Boggs." Rise Up for Asian Americans and Pacific Islanders - Spotlight at Stanford, exhibits.stanford.edu/riseup/feature/grace-lee-boggs.
- Grover, Rohan, and Rachel Kuo. "Destabilizing Race in Political Communication: Social Movements as Sites of Political Imagination." *Race and Ethnicity as Foundational Forces in Political Communication*, 7 Apr. 2024, pp. 118-137.
- Jubilee. "Are We Allies? Black Americans vs Asian Americans | Middle Ground." YouTube, 17 Jan. 2021, www.youtube.com/watch?v=pXo2ub_nZFc.
- Kim, Claire J. *Asian Americans in an Anti-Black World*. 2023.
- KIM, CLAIRE J. "The Racial Triangulation of Asian Americans." *Politics & Society*, vol. 27, no. 1, 1999, pp. 105-138.
- Kim, Claire J. "Asian Americans and Anti-Blackness." *Politics, Groups, and Identities*, vol. 10, no. 3, 2022, pp. 503-510, www.tandfonline.com/doi/full/10.1080/21565503.2021.2016448.
- Kuo, Rachel, and Sarah J. Jackson. "The political uses of memory: Instagram and Black-Asian solidarities." *Media, Culture & Society*, vol. 46, no. 1, 26 July 2023, pp. 164-186.
- Liu, Wen. "Black Lives Matter and Asian American Political Fracture." *Feeling Asian American: Racial Flexibility Between Assimilation and Oppression*, 2024, pp. 81-109. "One Moment, Please..." One Moment, Please..., freedom-inc.org/index.php?page=about-us.
- Ruiz, Neil G., et al. "3. Asian Americans and the 'model Minority' Stereotype." Pew Research Center, 2 Oct. 2024, www.pewresearch.org/race-and-ethnicity/2023/11/30/asian-americans-and-the-model-minority-stereotype/.
- "Thích Nhất Hạnh and Dr. Martin Luther King, Jr.: Spiritual Brothers, Partners in Nonviolence." USF Blogs – Educational Blogs from Our Community, 12 Oct. 2021, usfblogs.usfca.edu/fierce-urgency/2021/10/12/thich-nhat-hanh-and-dr-martin-luther-king-jr-spiritual-brothers-partners-in-nonviolence/.
- Tu, Lucy. "'ABGs', Affirmative Action, and Anti-Blackness in the Asian American Community." *Harvard Political Review*, 23 Mar. 2021, harvardpolitics.com/asian-american-solidarity/#google_vignette.
- "Yuri Kochiyama (U.S. National Park Service)." NPS.gov (U.S. National Park Service), 17 June 2021, www.nps.gov/people/yuri-kochiyama.htm.
- "Yuri Kochiyama." Rise Up for Asian Americans and Pacific Islanders - Spotlight at Stanford, exhibits.stanford.edu/riseup/feature/yuri-kochiyama.

A Systematic Review of the COVID-19 Lockdown Impacting Children with Autism: Changes in Applied Behavioral Analysis and Quality of Life

By Adriana Victoria

Abstract

The COVID-19 pandemic created significant uncertainty for many populations, particularly for children with Autism Spectrum Disorder (ASD) who rely on consistent, structured interventions like Applied Behavioral Analysis (ABA) therapy. This literature review examines how pandemic-related lockdowns disrupted access to ABA therapy and negatively affected both autistic children and their families. This study examines the impact of pandemic disruptions on therapy, diagnosis, and quality of life. I reviewed 67 of peer-reviewed studies identified through Google Scholar, JSTOR, and PubMed with the keywords mentioned below. Using combinations of these keywords and limiting results to articles published between 2020 and 2024, studies were included if they focused on child populations and excluded if they addressed only adults. Over 60% of children ($n = 3,502$; $p < .05$) experienced interruptions in ABA therapy, leading to halted developmental progress and regression in social and communication skills(White et al.). Additionally, delays in autism diagnoses were reported due to the lack of in-person assessments. Most families in reviewed studies reported heightened stress levels and decreased overall well-being during the lockdown period. These findings underscore the urgent need for adaptable therapeutic models and stronger telehealth infrastructure to maintain continuity of care during public health emergencies. By highlighting the challenges faced during the COVID-19 pandemic, this review contributes to a better understanding of how to support neurodivergent children in future crises.

Keywords

Autism Spectrum Disorder, COVID-19 Pandemic, Applied Behavioral Analysis, Neurodivergent children, Quality of Life

Introduction

According to the World Health Organization, the COVID-19 outbreak started in December 2019 in Wuhan, China (Alwine et al.). COVID-19 originally started as pneumonia cases related to the Huanan Seafood Wholesale Market; the problem arose when standard pneumonia treatments proved ineffective (Alwine et al.). Additionally, the inoculation of respiratory cells in the case showed similarities to the SARS-CoV, hence naming this novel virus SARS-CoV-2, more commonly known as COVID-19 (Wu et al.). In March 2020, the World Health Organization declared the novel coronavirus outbreak a global pandemic (Alwine et al.). It was hard to keep the virus under control because COVID-19 could travel in the air, infecting people around it. Hospital staff and beds were limited while cases increased, and there was no vaccine (Alwine et al.). The closure of schools and workplaces led to a widespread reliance on online platforms as the primary mode of communication (Cheng et al.). According to the World

Health Organization, between January 1, 2020, and December 31, 2021, almost 15 million people died because of cases directly or indirectly related to COVID-19. The COVID-19 pandemic led to the shutdown of public areas and confined people to their homes for over a year.

Autism Spectrum Disorder (ASD) is characterized by social and communication challenges and restrictive, repetitive behaviors. The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) classifies ASD as a developmental disorder because the initial symptoms appear in the first two years. It also states the criteria needed to be diagnosed with ASD, such as lack of communication and interaction, repetitive behaviors, and restricted interests (Lord). Some other common symptoms of ASD include avoiding eye contact and sensitivity to loud noises (Hyman et al.). People with ASD often avoid social interactions and public areas, leading them to have a lot of social anxiety, which is a fear of confronting people in social situations. One thing to note is that ASD is based on a spectrum, which means that some autistic people might have a higher sensitivity to stimuli than others. It is important to note that everyone's symptoms vary on different levels, so not every autistic person will respond or react the same way (Milton).

Social anxiety disorder, or social anxiety in short, is when one has an intense fear of social settings because they anticipate a negative outcome from the interaction (Leichsenring et al.). The research paper also claims that people diagnosed with this disorder are also scared about making mistakes, seeming incompetent, and fearing what other people may think about them. People with ASD are also prone to developing this disorder, which further diminishes their self-confidence.

Applied Behavioral Analysis (ABA) therapy is a personalized, in-person instruction for people with ASD and other neurological developmental disorders (Smith et al.). In the mid-1960s, Dr. Lovaas created a behavioral program that taught language to children with ASD. He made a program with the goal not to institutionalize children but rather to create a safe environment for children with disorders. These institutions were typically isolated from the public, reflecting societal misconceptions about autism and its causes. Professionals commonly held the view that "refrigerator mothers" were responsible for the symptoms observed in these children, attributing deficits in functioning to poor attachment and absentee parenting (Kanner et al.). ABA therapy emerged as an alternative to institutionalization and focused on integrating children with ASD into mainstream society. Therapies like ABA allow children to learn about safety in the world, understand how to act in social situations, and help them overcome their social anxiety. The therapy is designed to take place in school and public areas so that the children get more exposure to the outside world. The duration of ABA therapy depends on each child's needs, but it has an average of 3 to 5 years for 15 to 20 hours a week. While ABA therapy is beneficial to most patients facing ASD, it isn't the "perfect" treatment because each autistic patient is unique and may need help in different aspects of their lives.

One significant concern is that early iterations of ABA employed punitive measures to suppress undesirable behaviors. Although contemporary practices have largely abandoned such methods, critics argue that the therapy remains overly repetitive and may not adequately consider the child's emotional well-being (Child Mind Institute). Additionally, some autistic

self-advocates contend that ABA therapy can be detrimental by discouraging behaviors intrinsic to the autistic experience, such as stimming, which often serves as a coping mechanism. Suppressing these behaviors may lead to increased stress and emotional distress (Geng et al.). Critics also highlight that ABA's focus on modifying behaviors to align with neurotypical norms may inadvertently encourage masking, where autistic individuals suppress their natural behaviors to fit in. This can lead to negative mental health outcomes, including anxiety and identity confusion (Geng et al.).

When the COVID-19 pandemic was set in place, public areas were closed to prevent the spread of the virus, and this inevitably affected ABA therapy as well. Many therapy centers closed down, and only a handful provided online services. Social interaction decreased significantly, and everyone resorted to staying in their homes. ASD patients need a comfortable and predictable environment to stay calm and relaxed. With the sudden change of being confined to their homes and not getting the treatment that they needed, many families and children with ASD reported feeling isolated and lacking access to needed services. Face-to-face interactions are crucial for children with autism spectrum disorder (ASD) as they facilitate the development of social communication skills, including understanding non-verbal cues and engaging in reciprocal conversations. In-person interactions provide immediate feedback and support, essential for reinforcing appropriate behaviors and social norms (National Autism Center). Moreover, while beneficial in some contexts, the shift to remote learning and telehealth services often lacked the personalized support and social interaction that in-person settings provide. While prior research has documented the broader effects of COVID-19 on child mental health and the importance of ABA in supporting developmental outcomes, less is known about how pandemic-related disruptions specifically affected the accessibility and efficacy of ABA therapy for children with ASD.

Methodology

This review draws on peer-reviewed research articles published between January 2020 and March 2024 that explore the intersection of ASD, children, the COVID-19 pandemic, social anxiety, and ABA therapy. To identify relevant sources, I searched Google Scholar, JSTOR, and PubMed, three databases selected for their coverage of medical, psychological, and interdisciplinary literature.

Search Strategy

Specific keywords guided the search process to ensure comprehensive coverage. Keywords included: “*Autism Spectrum Disorder*” OR “*ASD*”, “*children*” OR “*pediatric populations*”, “*COVID-19 pandemic*” OR “*coronavirus*”, “*social anxiety*” OR “*mental health*”, and “*applied behavior analysis*” OR “*ABA therapy*” OR “*behavioral interventions*.” Example search strings were:

- "Autism Spectrum Disorder" AND children AND "COVID-19 pandemic" AND "applied behavior analysis."
- "ASD" AND "COVID-19" AND "social anxiety" AND "therapy"
- "children with autism" AND "COVID-19" AND "ABA"

Searches were limited to peer-reviewed journal articles. While JSTOR and PubMed inherently index peer-reviewed work, Google Scholar sometimes lists non-peer-reviewed content such as conference papers, reports, or preprints. To confirm peer-review status, I checked each Google Scholar result against the publisher's website, verified journal policies, and cross-referenced with PubMed or JSTOR when available. Any source that could not be verified as peer-reviewed was excluded.

Screening and Inclusion Process

The initial search yielded 212 records across the three databases. Titles and abstracts were reviewed, and studies were excluded if they:

1. Focused solely on adult populations rather than children or adolescents.
2. Did not directly address the COVID-19 pandemic.
3. Not published in peer-reviewed journals.
4. Lacked empirical data or methodological transparency.

Common exclusion reasons at the full-text stage included insufficient data, methodological flaws, or absence of relevance to ASD interventions during COVID-19. A simplified PRISMA-style flow was used to guide this process:

- Records identified: 212
- Records after duplicates removed: 164
- Records screened by title/abstract: 164
- Records excluded: 96
- Full-text articles assessed: 67
- Full-text articles excluded (with reasons): 33
- Final studies included: 34

This structured approach increases transparency and allows the review to be more replicable.

Quality Appraisal

To minimize subjective bias, I applied a structured checklist when evaluating the quality of included studies. Each study was reviewed against the following criteria:

- Study design (cross-sectional, longitudinal, experimental, or qualitative).
- Sample size and representativeness.
- Transparency of methods (clear reporting of measures, data collection, and analysis procedures).
- Statistical validity (appropriate methods, reporting of effect sizes, significance levels, or confidence intervals).
- Relevance (direct connection to children with ASD, COVID-19 disruptions, and/or ABA or related therapies).

Studies that met most criteria were prioritized, while those with significant methodological weaknesses or insufficient transparency were excluded. This checklist approach introduced consistency across the evaluation process, thereby reducing the influence of subjective judgment.

Data Extraction and Synthesis

Quantitative findings were extracted from tables, figures, and graphs, particularly those documenting disruptions in therapy access, variations in therapy effectiveness, or reported increases in social anxiety among children with ASD during the pandemic. For additional context, I also consulted authoritative resources such as the National Institutes of Health (NIH) and recognized autism research organizations.

To facilitate synthesis, studies were organized thematically according to both intervention type (e.g., ABA therapy, speech and language therapy, telehealth models) and observed outcomes (e.g., regression in communication, increased hyperactivity, heightened anxiety). This thematic categorization aligns with the structure of the Results and Discussion sections, ensuring consistency between methodology and interpretation.

Results

Therapy Disruptions During COVID-19

Based on numerous studies, many autistic patients had an increase in social anxiety. Although some treatment places continued online, over 60% of people ($n = 3,502$; $p < 0.05$) reported that there were disruptions in ABA therapy and speech and language therapy due to the COVID-19 pandemic (White et al.). According to the same study, the interruption of these crucial therapies led to an increase in behavioral challenges and a decrease in mental well-being. After the pandemic, many children with autism became more inactive and started spending more time with screens and electronic devices (Paterson et al.).

Recent research highlights several key areas of concern regarding ABA therapy during the pandemic. A study analyzing ABA therapy utilization in California from March 2019 to February 2021 found a notable decrease in therapy hours during the pandemic (Gonzales et al.). Specifically, there was an average reduction of 10.65 hours per month in the initial three months

following the pandemic's onset. Although there was a gradual increase of 2.38 hours per month over the subsequent nine months, therapy utilization had not returned to pre-pandemic levels even after a full year (Gonzales et al.). This disruption was more pronounced among children from Asian, non-Latinx backgrounds, and those funded by school districts, indicating disparities based on race, ethnicity, and funding sources (Amorim et al.).

Educational and Support Services

In addition to therapy centers, other services that were significantly impacted include educational programs and support services, which moved from in-person to virtual delivery during the pandemic (Zhou et al.). The closure of schools and extracurricular programs limited children's socialization opportunities, further increasing isolation and hindering the development of essential life skills (Vasilenko et al.). These disruptions not only affected the children's progress but also placed additional stress on families, who had to navigate these challenges without the usual support systems.

Behavioral and Psychological Outcomes in Children

The majority of caregivers, children, and adults with ASD said that the COVID-19 pandemic led to an increase in many negative emotions such as fear, nervousness, sadness, and guilt (Lugo-Marín et al.). In the long term, the pandemic was predicted to contribute to effects such as post-traumatic stress disorder and depression, and many children started to face sleep disorders (Lugo-Marín et al.). Another long-term effect was that many children started getting misdiagnosed. Since babies couldn't get their usual medical checkups, pediatricians were not able to diagnose babies with ASD, leading to further delays in babies getting ASD treatment (Bellomo et al.). After the pandemic, social anxiety increased among children; as a result, many parents mistake these symptoms for normal behavior, leading to missed ASD diagnoses and delaying access to needed therapy (Bellomo et al.). Many children, who were unable to express their frustration in words, often show their distress through tantrums and aggression (Bellomo et al.).

Caregiver and Family Stress

The pandemic's impact extended beyond therapy disruptions, affecting children's behaviors and family dynamics. A survey revealed that over 74% of caregivers observed moderate-to-large negative changes in their children's behaviors during the pandemic (Panjwani et al.). The same study also stated that families experiencing food insecurity and those with annual incomes below \$50K reported more significant behavioral challenges. The pandemic also disrupted established routines, which are particularly important for children with ASD, leading to increased anxiety and frustration (Jones et al.). Most families in reviewed studies reported heightened levels of stress, with many observing a decline in their children's social skills, communication, and emotional regulation, which are vital for their overall development (Masi et al.). Common issues included increased distractibility, hyperactivity, and sleep disturbances,

which are associated with co-occurring conditions like attention deficit hyperactivity disorder (ADHD) (Baweja et al.).

Elevated stress levels among parents during the pandemic further influenced the family's quality of life. In research studies, the *Quality of Life* is defined with the L-QoL, which refers to an individual's overall well-being, encompassing physical, psychological, and social aspects of life, as well as their satisfaction with these domains. It is a subjective, multidimensional concept that varies based on personal values, cultural influences, and life circumstances. Research indicates that pandemic-related stress adversely affects family dynamics, with parental involvement playing a mediating role (Cheng et al.). Families with higher stress levels and lower parental involvement reported a diminished quality of life in several key areas, including a decrease in family cohesion, emotional well-being, and the ability to effectively manage daily routines (Cheng et al.). The stress from managing challenging behaviors and the lack of external support led to increased anxiety, depression, and overall dissatisfaction with family life. Additionally, the limited access to therapy and educational resources further hindered the children's social and developmental progress, making it more difficult for parents to balance caregiving responsibilities with personal and professional commitments. These findings suggest the need for policies that support parental engagement and address stressors to improve family well-being (Colizzi et al.).

Healthcare Access and Alternative Delivery Methods

The pandemic also affected access to essential healthcare services. Approximately one-third of children with autism and other special healthcare needs missed preventive checkups due to the pandemic, and about half utilized virtual care services in 2021 (Rast et al.). This shift underscores the challenges families face in maintaining routine healthcare, which is crucial for managing ASD (L. A. Vismara & S. J. Rogers). In response to the challenges posed by the COVID-19 pandemic, alternative service delivery methods were implemented to support children with autism spectrum disorder (ASD). Telemedicine emerged as a viable option, with studies demonstrating its effectiveness in providing direct interventions and parent-mediated programs aimed at enhancing children's adaptive skills (Smith et al.). Additionally, online parent-mediated home-based interventions were utilized, enabling parents to actively engage in their child's therapy under professional guidance, thereby maintaining developmental progress during periods of restricted access to in-person services (Brown et al.).

Discussion

The findings of this review align with a growing body of literature documenting the widespread disruptions experienced by children with Autism Spectrum Disorder (ASD) during the COVID-19 pandemic. Consistent with prior studies, the closure of therapy centers and schools led to regression in communication skills, heightened anxiety, and the emergence of aggressive behaviors in the home environment. These outcomes mirror trends observed in broader pediatric populations but were particularly pronounced for children with ASD due to

their reliance on routine, structured interventions, and consistent social engagement. Other reviews have similarly reported that therapy interruptions contributed to long-term setbacks in developmental milestones, suggesting that children with ASD were disproportionately vulnerable to the sudden changes brought about by the pandemic.

Beyond disruptions in therapy, the results highlight disparities in access that were shaped by socioeconomic and racial factors. Studies showing greater declines in ABA utilization among children from Asian, non-Latinx backgrounds and school districts—funded services point to systemic inequities that limited families’ ability to maintain care. This is consistent with pre-pandemic research indicating that children from minority or low-income families already faced barriers in accessing ASD interventions. The pandemic appears to have amplified these disparities, widening existing gaps in developmental support and contributing to differential outcomes across populations.

The review also underscores the interdependence between children’s outcomes and family well-being. Elevated caregiver stress, reduced family cohesion, and declines in quality of life were common across studies. This finding echoes prior literature on the stress-coping model, which emphasizes how parental mental health directly mediates child outcomes in ASD populations. The inability to rely on schools, therapy centers, or extended support networks forced families to shoulder additional responsibilities, often with limited resources. These results suggest that interventions aimed at supporting children must also address caregiver resilience and family systems if they are to be effective in the long term.

Conclusion

When the COVID-19 pandemic took place, many therapy centers closed down, leading to many families being confined in their homes. Individuals with ASD experienced restricted access to external stimuli, which contributed to heightened feelings of confinement within their homes. During the pandemic, children with ASD exhibited increased aggression in the home environment, and even after restrictions were lifted, many did not return to their previous behavioral baseline. They weren’t able to have conversations without being anxious. The progress that they made at ABA therapy was lost, their conditions worsened, and the quality of life for families decreased due to increased stress levels. The long-term consequences of these disruptions could be profound, affecting both the children with ASD and their families. Prolonged social isolation and the lack of structured therapy may lead to significant setbacks in social and communication skills, making it harder for children to reintegrate into social settings like school and community activities (Kuhn & Hensley). Additionally, the rise in anxiety and aggressive behaviors could contribute to the development of more severe mental health issues, such as depression or severe emotional regulation difficulties (Sinha et al.). Over time, the lack of early intervention and routine may lead to a need for more intensive support in adulthood, as children may struggle to manage their daily lives independently. As a result, families may experience increased financial and emotional strain, which could affect their overall well-being and the long-term outcomes for both the child and caregivers.

The pandemic significantly affected many children with autism, exacerbating social anxiety and creating additional challenges for both the children and their caregivers. Sudden changes in routine, common during the pandemic, can be particularly stressful for children with ASD, as they may struggle to adapt (Kuhn & Hensley). Moreover, since verbal communication is often limited in children with ASD, they may resort to aggressive behaviors to express their distress (Sinha et al.). These behaviors, if left unaddressed, can lead to long-term mental health issues such as depression and anxiety (Vasilenko et al.).

To mitigate these impacts, it is crucial to implement strategies that support emotional expression and reduce stress. Caregivers and educators should focus on providing stable routines, creating safe spaces for children to express their emotions, and seeking professional guidance when needed (Zhou et al.).

Limitations

This review has several limitations that should be considered when interpreting the findings. First, the analysis is based on 34 published, peer-reviewed articles available between January 2020 and March 2024. While this ensures timeliness, it also means that more recent or ongoing studies may not have been captured, and the full long-term consequences of COVID-19 on children with ASD remain uncertain.

Second, sample size and study design varied widely across included studies. Some relied on large caregiver surveys with thousands of participants, while others were based on small regional cohorts. This heterogeneity reduces the ability to generalize findings and complicates direct comparisons. Many studies also relied on self-report surveys, which may introduce recall bias or subjective interpretation of children's behaviors.

Another aspect that could be explored more deeply is what other aspects could be correlated with ASD. Another issue was that people were required to wear masks once ABA therapy centers started opening up. Some children are sensitive to stimuli and struggle to wear masks, so how did different therapy centers address this issue? The increase in social anxiety has been related to other disorders like depression.

Works Cited

1. M. Worobey, J. I. Levy, L. Malpica Serrano, A. Crits-Christoph, J. E. Pekar, S. A. Goldstein, A. L. Rasmusen, M. U. G. Kraemer, C. Newman, M. P. G. Koopmans, M. A. Suchard, J. O. Werthem, P. Lemey, D. L. Robertson, R. F. Garry, E. C. Holmes, A. Rambaut & K. G. Andersen. The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. *Science*, 377(6609), 951–959 (2022).
2. Y. Wu, W. Ho, Y. Huang, D. Y. Jin, S. Li, S. L. Liu, X. Liu, J. Qiu, Y. Sang, Q. Wang, K. Yuen, & Z. M. Zheng. SARS-CoV-2 is an appropriate name for the new coronavirus. *The Lancet*, 395(10228), 949–950 (2020).
3. S. Cheng, Y. Yang, & M. Deng. Psychological stress and perceived school success among parents of children with developmental disabilities during the COVID-19 pandemic. *Journal of Autism and Developmental Disorders*, 52(7), 3194–3201 (2022).
4. C. Lord. *Autism Spectrum Disorder: A Comprehensive Guide*. Springer (2018).
5. S. L. Hyman, S. E. Levy, S. M. Myers, D. Z. Kuo, S. Apkon, L. F. Davidson, K. A. Ellerbeck, J. E. A. Foster, G. H. Noritz, M. Leppert, B. S. Saunders, C. Stille, L. Yin, C. C. Weitzman, D. O. Childers, J. M. Levine, A. M. Peralta-Carcelen, J. K. Poon, P. J. Smith, N. J. Blum, J. I. Takayama, R. Baum, R. G. Voigt & C. Bridgemohan. Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics*, 145(1) (2020).
6. D. Milton. On the ontological status of autism: The ‘double empathy problem’. *Disability & Society*, 27(6), 883–887 (2012).
7. E. Courchesne, K. Pierce, C. M. Schumann, E. Redcay, J. A. Buckwalter, D. P. Kennedy, & J. Morgan. Mapping early brain development in autism. *Neuron*, 56(2), 399–413 (2007).
8. M. J. Maenner et al. Prevalence of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2016. *MMWR Surveillance Summaries*, 69(4), 1–12 (2020).
9. F. Leichsenring & F. Leweke. Social anxiety disorder. *New England Journal of Medicine*, 376(23), 2255–2264 (2017).
10. T. Smith & S. Eikeseth. O. Ivar Lovaas: Pioneer of applied behavior analysis and intervention for children with autism. *Journal of Autism and Developmental Disorders*, 41, 375–378 (2011).
11. L. Kanner. Autistic disturbances of affective contact. *Nervous Child*, 2, 217–250 (1943).
12. D. N. Stern. The psychoanalytic treatment of autistic children. *The Psychoanalytic Quarterly*, 29(1), 51–74 (1960).
13. Child Mind Institute. The controversy around ABA. (2024).
14. C. Geng. Is ABA therapy harmful? The controversy explained. *Medical News Today* (2024, November 27).
15. National Autism Center at May Institute. Historical perspective. (2024).
16. L. C. White, J. K. Law, A. M. Daniels, J. Toroney, B. Vernoia, S. Xiao, P. Feliciano, W. K. Chung, & SPARK Consortium. Brief report: Impact of COVID-19 on individuals

- with ASD and their caregivers: A perspective from the SPARK cohort. *Journal of Autism and Developmental Disorders*, 51(10), 3766–3773 (2021).
17. D. C. Paterson, K. Ramage, S. A. Moore, N. Riazi, M. S. Tremblay, & G. Faulkner. Exploring the impact of COVID-19 on the movement behaviors of children and youth: A scoping review of evidence after the first year. *Journal of Sport and Health Science*, 10(6), 675–689 (2021).
 18. J. Lugo-Marín, L. Gisbert-Gustemps, I. Setien-Ramos, G. Español-Martín, P. Ibañez-Jimenez, M. Forner-Puntonet, G. Artetaga-Henriquez, A. Soriono-Dia, J. D. Duque-Yemail & J. A. Ramos-Quiroga. COVID-19 pandemic effects in people with autism spectrum disorder and their caregivers: Evaluation of social distancing and lockdown impact on mental health and general status. *Research in Autism Spectrum Disorders*, 83, 101757 (2021).
 19. L. Zhou, J. Zhang, & L. Zhang. Strategies for supporting children with ASD during the COVID-19 pandemic. *International Journal of Developmental Disabilities*, 67(3), 202–210 (2021).
 20. S. A. Vasilenko, R. F. Leeman, & H. A. Tindle. Long-term mental health outcomes of children with ASD: Depression and anxiety post-pandemic. *Psychological Medicine*, 50(11), 1824–1832 (2020).
 21. T. R. Bellomo, S. Prasad, T. Munzer, & N. Laventhal. The impact of the COVID-19 pandemic on children with autism spectrum disorders. *Journal of Pediatric Rehabilitation Medicine*, 13(3), 349–354 (2020).
 22. C. W. Gonzales, J. R. Simonell, M. H. Lai, S. R. Lopez, & J. Tarbox. The impacts of the COVID-19 pandemic on therapy utilization among racially/ethnically and socio-economically diverse autistic children. *Journal of Autism and Developmental Disorders*, 53(3), 918–933 (2023).
 23. R. Amorim, S. Catarino, P. Miragaia, C. Ferreras, V. Viana, & M. Guardiano. The impact of COVID-19 on children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*. (2020).
 24. A. A. Panjwani, R. L. Bailey, & B. L. Kelleher. COVID-19 and behaviors in children with autism spectrum disorder: Disparities by income and food security status. *Research in Developmental Disabilities*, 115, 104002 (2021).
 25. C. Jones, A. Ellis, & H. Williams. The effect of COVID-19 lockdown on children with autism spectrum disorder: A qualitative study. *Journal of Developmental and Behavioral Pediatrics*, 42(4), 287–296 (2021).
 26. A. Masi, N. Glozier, R. C. Dale, & A. J. Guastella. The role of the immune system in the development of autism spectrum disorder. *Neuropsychopharmacology*, 46(1), 1–12 (2021).
 27. R. Baweja, S. L. Brown, E. M. Edwards, et al. COVID-19 pandemic and impact on patients with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 52, 473–482 (2022).

28. S. Cheng, S. Cheng, S. Liu, & Y. Li. Parents' pandemic stress, parental involvement, and family quality of life for children with autism. *Frontiers in Public Health*, 10, 1061796 (2022).
29. M. Colizzi, E. Sironi, F. Antonini, M. L. Ciceri, C. Bovo, & L. Zoccante. Psychosocial and behavioral impact of COVID-19 in autism spectrum disorder: An online parent survey. *Brain Sciences*, 10(6), 341 (2020).
30. J. E. Rast, K. H. Koffer Miller, J. Bromberg, J. Ventimiglia, K. A. Anderson, & L. L. Shea. Changes in child health care, health, and caregiver mental health during the COVID-19 pandemic in children with autism and special health care needs. *Maternal and Child Health Journal*, 1–9 (2024).
31. L. A. Vismara & S. J. Rogers. Behavioral treatments in autism spectrum disorder: What do we know? *Annual Review of Clinical Psychology*, 6, 447–468 (2010).
32. A. C. Smith, E. Thomas, C. L. Snoswell, H. Haydon, A. Mehrotra, J. Clemensen, & L. J. Caffery. Telehealth for global emergencies: Implications for coronavirus disease 2019 (COVID-19). *Journal of Telemedicine and Telecare*, 27(5), 309–313 (2021).
33. J. L. Brown, C. J. Smith, & A. M. Jones. Online parent-mediated home-based intervention for children with autism spectrum disorder during the COVID-19 pandemic: A pilot study. *Journal of Autism and Developmental Disorders*, 52(4), 1234–1245 (2022).
34. K. L. Kuhn & D. Hensley. The impact of the COVID-19 pandemic on children with autism and their caregivers. *Journal of Autism and Developmental Disorders*, 51(4), 1224–1235 (2021).
35. S. Sinha, R. Hedges, & K. Patel. Aggressive behaviors and communication challenges in children with autism spectrum disorder during the pandemic. *Journal of Child Psychology*, 63(2), 251–266 (2022).

Economic Impact of the VASEN 2025 Linhai Youxi Ultra Trail Triathlon on Taizhou, Zhejiang Province: A Comparative Analysis with the Tsaigu Trail Race Series

By Xinle (Simon) Chen

1. Introduction

In April 2025, the Linhai government, in collaboration with the Shanghai Triathlon Club (STC) Events, organized one of the most unique and challenging triathlon races in China: the VASEN 2025 Linhai Youxi Ultra Trail Triathlon. With strong sponsorship and increasing visibility as part of Zhejiang Province's sports-driven rural development initiative, the organizers are now able to evaluate the broader impact of the event on Youxi Town, Linhai City, and the surrounding region. The following sections will explore the event's operational footprint, assess the contributions from athlete participation and visitor spending, and conclude with an economic model that highlights the triathlon's influence on regional growth. Additionally, this report will compare its scope and impact with that of the Tsaigu Trail Race Series and offer my own policy recommendations.

2. Background and Context

2.1 The Youxi Triathlon Race

The Linhai Youxi Ultra Trail Triathlon is being held for the third time since 2023 and has attracted hundreds of athletes from over 18 different countries and regions, along with their families, spectators, volunteers, and operational staff (People's Daily Zhejiang). Every year, the race serves as an excellent platform to showcase Linhai's natural beauty and cultural heritage while providing exciting opportunities for local businesses and tourism operators.

2.2 The Tsaigu Race: A Benchmark for Success

The Tiantai County Government and the Tsaigu Organizing Committee host one of China's most remarkable and internationally recognized trail running events: the Tsaigu Trail Running Festival, which has been held annually since 2015. This event attracts thousands of trail runners from over 20 countries and regions, along with supporters, spectators, volunteers, and event staff. Set against the stunning backdrop of Tiantai Mountain's rugged terrain and cultural landmarks, the festival serves as an exciting platform to showcase the area's natural beauty and Taoist heritage while creating significant opportunities for local tourism and small businesses (Xinhua News). With its growing reputation as a leading international endurance event and its inclusion in global trail running circuits, the Tsaigu Festival is increasingly seen as a model for rural sports tourism in Zhejiang Province.

3. Methodology

To ensure accuracy and credibility in evaluating the economic impact of the VASEN 2025 Linhai Youxi Ultra Trail Triathlon and comparing it with the Tsaigu Race, data were

collected and analyzed using both primary and secondary sources, which I gathered or obtained from online sources cited at the end of this report.

First, a questionnaire was distributed through the Taizhou Triathlon's participant group chat to gather quantitative data on spending and visitor experiences. A total of 29 valid responses were received, covering accommodation costs, transportation fees, dining expenses, event participation, and overall satisfaction. The questionnaire was effective in helping me conduct a quantitative analysis and predict the race's economic impact.

Second, qualitative insights were obtained through in-depth interviews. I interviewed 5 athletes, 3 local vendors, 2 sponsors, 2 government officials, and 2 event volunteers during and after the event. The questions focused on participants' overall experiences, their intentions to participate in the event, differences in customer numbers, the overall goals of the race, and suggestions for local hospitality and facilities. These interviews provided perspectives that complemented the quantitative survey data and helped me better understand the impact on the local community.

Third, the total economic impact was calculated by multiplying the average total spending per visitor by the total number of visitors, including both athletes and accompanying spectators. This calculation represents the average total spending per visitor across all categories included in my survey.

Finally, there are several limitations that should be acknowledged. The questionnaire data relied on voluntary participation, which resulted in a limited sample size. Additionally, primary economic data for the Tsaiyu Race were not directly available, and the secondary sources were drawn from media reports and official government announcements.

4. Economic Impact Analysis of the Youxi Triathlon Race

4.1 Athlete Spending: Preparation and Participation

According to my survey (Figure 1), around 83% of participants stayed in the local area for 2 to 3 days. Most participants were not local residents, so a majority (64%) reported spending between 100 and 500 CNY on accommodation. About 50% of the participants spent a similar amount on travel fees, while 75% reported spending between 100 and 500 CNY on food. Additionally, participants contributed further to the local economy by spending on activities outside the triathlon events, as the registration fee for the triathlon itself was only 800 CNY. The most popular activity among participants was dining out; up to 75% spent between 100 and 500 CNY on local food (Survey). This not only benefits local restaurants but also introduces participants to more attractions, increasing the likelihood that they will stay longer and spend more on food and travel during future visits.

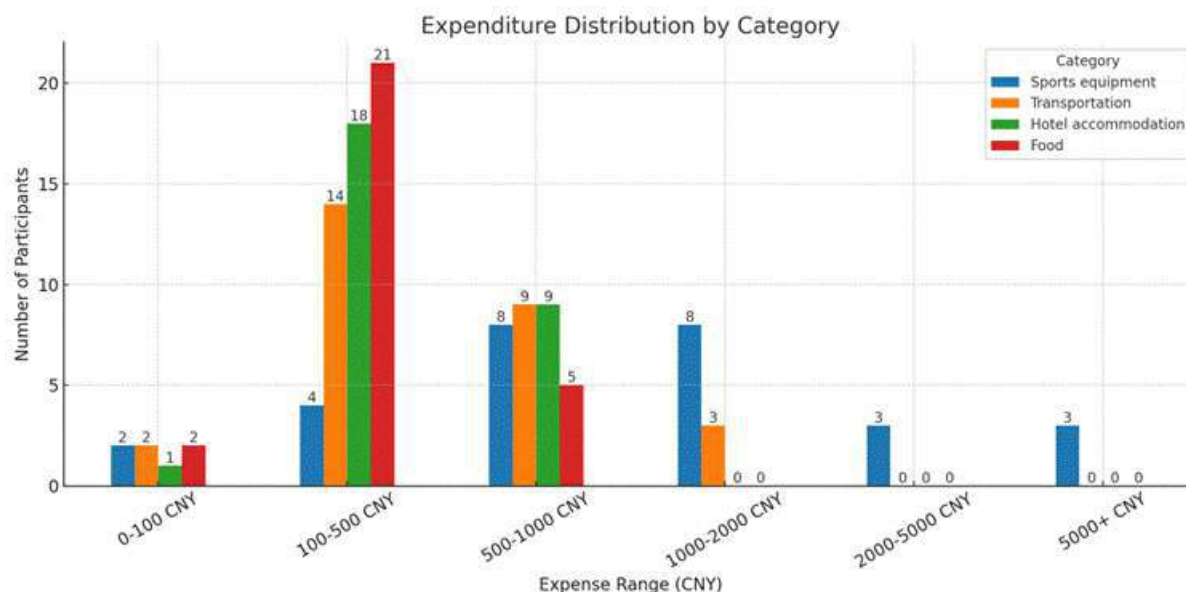


Figure 1: Athlete survey responses on expenditures

4.2 Local Business and Tourism Impact

According to official government reports, Youxi expects around 50,000 visitors to the city next year after hosting the event. Moreover, the city has already hosted over a hundred high-quality sports events throughout the year, which further boosts the growth of industries related to sports goods and services in Linhai. Nearly 160 sports enterprises in Linhai have been encouraged to expand their efforts in event organization, service provision, and equipment manufacturing. As a result, the total output value of the sports industry has exceeded 10 billion CNY, creating more than 100,000 jobs (People's Daily Zhejiang).

4.3 Infrastructure and Long-Term Benefits

At the same time, the local government can maintain the cross-country track to promote racing, academic research, and hiking events, aiming to make the city a year-round destination without an off-season. This approach can not only enhance accommodation facilities and expand restaurant services but also provide greater exposure for local agricultural products. Ultimately, it will bring new vitality to the local economy and attract more visitors in the long term.

To illustrate the economic ripple effects of sports events on regions similar in size to Linhai Youxi, let's consider Anji, a county-level region in Zhejiang. I make this comparison because the Linhai government has not provided direct data, so I can only predict possible outcomes in this way. Anji hosts the annual National Mountain Bike Open, which attracts 2,000 national participants. Through local and national television broadcasts and media promotion, the event receives over 230 million media impressions annually, leading to significant exposure and recognition (Sohu Sports).

During race weeks, Anji welcomes 80,000 tourists, with up to 70% coming from outside the region. Hotel occupancy rates increase from 50% to 95%, and homestay revenues rise by

65%. Local restaurants also benefit, increasing their revenue by 40%, while sales of regional specialties like bamboo shoots and tea soar by 200%. Additionally, 10 professional cycling trails have fostered industrial growth: equipment rental, outdoor training, and related services contribute 300 million CNY annually to the local economy. The sports industry’s contribution to Anji’s GDP has increased from 1.2% to 3.5%, creating 2,000 jobs (Sohu Sports).

This example shows how events like Linhai’s triathlon can boost tourism, brand building, and industrial upgrading. Modeled on Anji’s sports-driven growth, Youxi could experience significant economic gains from its triathlon events.

4.4 Predictions for Future Economic Growth Modeling on the Anji Example

Modeled after Anji, Youxi’s triathlon could lead to a significant boost in the local economy. In the short term, 15,000 visitors could bring in 15 million CNY in spending, which might increase homestay revenues by up to 65%. In the long term, annual GDP growth could rise by 0.8% to 1.2%, creating 1,200 direct jobs. Additionally, the tourism brand value could increase by 20%, while converting event traffic into sustained growth (Sohu Sports).

5. Comparative Analysis: Youxi Ultra Trail Triathlon vs. Tsaigu Trail Race Series

5.1 Scale and Participation

When comparing the Youxi triathlon with the Tsaigu race, both attract hundreds of athletes from over 18 different countries and regions, with around 2,000 participants in the Taizhou triathlon. This event has a tight schedule, taking place within a single day from approximately 7:00 AM to 5:00 PM, including swimming, biking, and running segments in sequence, with a maximum duration of 10 hours (Xinhua News).

In contrast, the Tsaigu race is a well-known trail running competition that usually sees 4,500 athletes sign up from a pool of 14,400 applicants, with participants from more than 30 countries. Its initial appeal was mostly domestic, but it has gained international recognition thanks to its Ultra-Trail du Mont-Blanc (UTMB) qualification status and growing fame through social media and trail running networks. The Tsaigu event lasts 2 to 3 days and features multiple race categories, such as 25K, 50K, and 105K. The longest race starts at 6:00 AM on Day 1 and finishes by noon on Day 3, allowing participants to pace themselves freely (Xinhua News).

The table below summarizes key differences between the two events in terms of participation, race format, sponsorship, international influence, and economic impact:

Dimension	2025 Linhai Youxi Extreme Triathlon	2024 Tsaigu Trail Running Race
Participants	~2,000 athletes, 18+ countries	~4,500 athletes, 30+ countries
Race Format	Triathlon (swim-bike-run), 1-day, 10-hour limit	Ultra trail (25K/50K/105K), 2-3 days
Sponsors	11 sponsors, title sponsor Weixing Group	Multiple sponsors, incl. Salomon
International Influence	Emerging, viral in China (~100M online views)	High, UTMB-qualifying, global elite runners
Economic Impact	~¥10M local revenue (tourism, F&B, hotels)	~¥30M direct, >¥100M indirect impact

Table 1 Comparison in scale and participation: Youxi Triathlon vs. Tsaigu Race

5.2 Community and Branding Effect

There are also differences in community engagement and branding between the two races. The STC Triathlon in Youxi involves 220 local volunteers, helping the community become more familiar with the sport. After the race concludes, the facilities are repurposed as a “Youth Triathlon Training Centre,” linking the event to long-term youth sports development.

From a branding perspective, the triathlon has the support of 11 sponsors, one of the largest being a local Linhai company that provides water during the race. This event positions Linhai as a “distinct place for relaxation, away from the cities,” building its reputation as an excellent outdoor getaway where people can escape work pressures and urban pollution. This may encourage more visitors to come to Linhai, even outside of the racing season (Xinhua News).

Meanwhile, the Tsaigu race is deeply connected with multiple rural communities, engaging over 300 villagers who serve as supply station staff and volunteers. This strengthens the bond between the race and the local population. Its branding as a “Mountain Trail Running Paradise” has been recognized by iRunFar as “China’s most culturally rich trail race,” thanks to its blend of untouched trails and immersive traditional cultural experiences (Zhejiang Women’s Federation).

Economically, the Tsaigu race generates roughly three times more direct revenue (about ¥30 million) and over ten times the indirect economic impact (greater than ¥100 million) compared to the Youxi triathlon's approximately ¥10 million local revenue (Xinhua News). This reflects Tsaigu’s larger scale, longer duration, and broader global reputation. Therefore, it could serve as a good role model for the Youxi triathlon to imitate in order to achieve a greater overall impact .

6. Athlete Perspectives and Experiences

I personally competed in this triathlon and was the only participant under 18 for two years in a row. Additionally, I conducted interviews with multiple people of various professions, including sponsors, vendors, participants, volunteers, organizers, local workers, and food vendors. I collected a significant amount of data from them, which is useful for further investigating the economic impact.

As a participant, I gained valuable insights from the experience. The first part of the race is swimming, and with intense music playing from large speakers, I jumped into the reservoir surrounded by mountains. The chilly water eased my nerves and helped me focus on every stroke and breath. The 1.5-kilometer swim was shorter than I expected, and soon I was on to the cycling segment. After getting out of the water and into the transition zone, a red carpet protected my feet as I ran to my bike. After putting away my swimming gear and putting on my shoes and helmet, I started cycling.

Cycling through the mountains on smooth, well-constructed asphalt roads was an extraordinary experience, even though I had to climb 1,000 meters. Moving on to the cross-country run, I parked my bike at the peak, changed my shoes, and grabbed my running gear. The 25-kilometer run featured 4 supply stations that offered emergency medical support, helped with my cramps, and provided food, including carbohydrates, fruit, and soup. These items were essential for meeting the athletes' needs. Crossing the finish line was a proud moment, especially with volunteers cheering us on. After finishing, I received a warm blanket and headed to the free food vendor, where I enjoyed hot soup, noodles, and baozi—all for free. It was an excellent personal experience.

Before and after the race, I interviewed people from various sectors and received positive feedback. Some significant differences I noticed that set this race apart include the thoughtful design of the event. First, in the Youxi triathlon, professional athletes were placed at the front during the start to prevent them from having to overtake slower participants, which could disrupt their competitive edge. Second, the intervals between the different segments were well arranged, improving the experience for all participants. The entry point for the swimming segment was set at a higher platform, allowing athletes to dive in more efficiently. Third, although the transition zone was a bit far from the swimming finish point, red carpets were laid out along the path to protect athletes' feet as they walk to the transition area. Finally, after finishing the race, local food provided by villagers was free. This not only fulfilled participants' basic needs but also introduced them to the local culinary culture, leaving a positive impression and potentially encouraging them to support local businesses.

However, the Youxi triathlon also had areas that needed improvement. For example, lifeboats that ensured swimmers' safety sometimes came too close, creating waves that can hinder performance. Additionally, there were imprecise signs showing the remaining distances during the cycling segment. While this issue might seem minor, incorrect signs could lead to confusion among participants.

Despite these challenges, the Youxi triathlon created an unforgettable experience with its vibrant atmosphere and thoughtful support system for participants. The final segment involved a cross-country run, which differs from other triathlon events. At steep locations or near wet streams, many professional volunteers from the mountaineering association were stationed to support participants and ensure their safety. Another impressive aspect was the large number of enthusiastic volunteers and villagers cheering participants as they passed by their homes. Their encouragement boosted my energy and motivation.

On the other hand, the basic facilities still needed improvement. For example, after completing the triathlon, which took about eight hours, participants were sweaty and exhausted, but the shower facilities were broken, and there was no hot water. This could cause discomfort or even illness, as participants could not take a proper shower. Therefore, I believe it is essential to improve and enhance the basic facilities to provide participants with a truly well-rounded experience.

7. Challenges and Limitations

While this analysis and comparison offer interesting findings and meaningful insights into the economic and social impacts of the Youxi Triathlon and the Tsaigu Race, several challenges and limitations must be acknowledged.

First, the economic impact predictions for the Youxi Triathlon relied primarily on data from my survey with limited responses from the participants in the official race group chat and a relatively small sample size, which may not fully reflect the diversity of participants' spending behaviors or capture long-term economic benefits. Meanwhile, the other source to investigate the Youxi Triathlon economy is interviewing local shop owners, which was recorded by camera; however, due to the limited responses from them, the result might not very effectively reflect the true situation either.

Second, meaningful comparisons between the Youxi Triathlon and the Tsaigu Trail Race remain constrained by differences in event scale, duration, and international exposure. While the Tsaigu Race has a longer history, broader global branding, and standardized official reporting on participation and revenue, the Youxi Triathlon is still developing its data-collection mechanisms to improve the accuracy of economic impact data.

Third, external factors such as weather conditions, transportation, and the timing of the event within the broader sports calendar may have influenced participant turnout, tourist spending, and the overall visitor experience in Youxi, introducing additional uncertainty into the economic estimates. For example, I visited Youxi for a triathlon twice in a row, and it rained each time. The rain forced me to take a shower after the race, so I ended up going to downtown Taizhou to check into an hourly room. This created extra expenses that would not have occurred if the weather had been better.

Finally, regional challenges—such as limited accommodating spots, insufficient marketing reach—such as participants who can't swim will not participate Youxi triathlon, whereas cross-country run would be suitable for most of the healthy people, and gaps in multilingual communication—may have limited the event's capacity to maximize its tourism and popularity among international communities. Solving these constraints well in the future would not only improve the precision of economic impact estimates but also strengthen Youxi's ability to compete on a larger scale, and become a role model for international sports tourism destinations.

8. Recommendations for Achieving Tsaigu-Level Impact

To help the Youxi Triathlon reach the scale and economic impact of the Tsaigu Race, I have proposed the following three strategies:

First, the organizer could strengthen international outreach by creating official event accounts on major global platforms like YouTube and Instagram. Regularly release professional videos and social media content featuring keywords such as "escape from the city," "outdoor lifestyle," and "relaxation." This will position the triathlon not just as a competition, but as a lifestyle experience that allows participants and visitors to reconnect with nature and escape

urban stress. The intended audience could be people living near Youxi, such as those in Shanghai, Suzhou, and Changshu.

Second, the local government and businesses could invest in high-quality accommodation facilities near the race area. Comfortable and affordable hotels would encourage participants and spectators to stay in Youxi longer, which would, in turn, boost local restaurants, shops, and entertainment businesses. This would multiply the economic benefits and enable long-term growth in tourism revenue.

Finally, training more professional volunteers and hosting cultural festivals alongside the race would help enhance community participation. These events can attract more people and generate interest in the local culture. Additionally, inviting social media influencers and using live-streaming platforms can also increase media exposure and promote the triathlon.

Works Cited

- Survey. 2025 Youxi Extreme Triathlon Internal Data: Participants, Sponsors, Economic Revenue, Volunteers, Branding, and Online Exposure. Event organizer, 2025.
- Xinhua News. "Small City, Big Event: Why Tsaigu Trail Running Is Booming." *Xinhuanet*, 12 Nov. 2024,
www.xinhuanet.com/20241112/76fb9532b5eb47e4b04345902749ba0c/c.html.
- Sohu Sports. "Tsaigu Trail Race: 4,500 Runners Gather at the Great Wall of Jiangnan, with Over 14,000 Pre-Registrations." *Sohu*, 22 Apr. 2025,
www.sohu.com/a/887635696_122362510.
- People's Daily Zhejiang. "9th Tsaigu Trail Race to Be Held in Linhai." *People.com.cn*, 25 Oct. 2024, zj.people.com.cn/BIG5/n2/2024/1025/c370990-41020396.html.
- Zhejiang Women's Federation. "Tsaigu Trail Race Drives Rural Revitalization." *Zjswomen.org.cn*, 21 Apr. 2025,
zjswomen.org.cn/nxfc/gflstll/202504/t20250421_23400152.html.

Moral Intuition Should Not Be Trusted By Xinyi Wu

Abstract

This essay argues that people have no good reason to trust moral intuition. First, within metaethical views, it posits moral facts. Subjectivism and relativism are rejected for undermining genuine moral disagreement and making morality implausibly malleable. Moral realism is then criticized using J. L. Mackie's queerness argument, that is, objective moral facts would be metaphysically strange and explanatorily unnecessary. Objections based on "companions in guilt" are answered by distinguishing truth-related reasons from action-guiding reasons. Finally, a skeptical regress argument shows that no moral belief can be justified, drawing on the idea that one cannot derive an "ought" from an "is," concluding that by lacking moral facts and justification, moral intuition should not be trusted.

Introduction

The concept of trust is intimately linked to the concepts of belief and truth. For if we ask whether we ought to trust someone or something (say a book or news report), what we are really asking is whether we ought to believe what we're being told, I.e., whether we ought to regard it as true. It is therefore natural to parse the question, "do we have any good reasons to trust our moral intuition?", as "do we have any good reasons to believe - take as true - that which moral intuition tells us?"

In this essay, I will argue that the answer is 'no' on the grounds that (1) there is no reason to believe in moral facts and so no reason to think that our beliefs about morality even *could* be true, and (2) there is no reason to think any particular moral belief is justified and so no reason to think our beliefs about morality are *actually* true. First, I will clarify my argumentative strategy and offer some considerations against metaethical theories that posit mind-dependent moral facts. Then, I will argue against moral realism. Finally, I will conclude with a skeptical regress argument against moral justification.

Metaethics

In contrast to normative ethics, which attempts to discover the principles governing morally right action and conduct, metaethics investigates the meaning of moral statements and the nature of moral reality (including the question of whether such a thing even exists). Metaethical theories admit of a broad division between realist accounts and non-realist accounts. A realist believes that the world contains objective, mind-independent moral facts. A non-realist is simply someone who denies realism.

On this definition, non-realism encompasses a wide and heterogenous range of views. Most non-realists will deny the existence of moral facts altogether, but some may accept them. Among the latter group are *subjectivists*, who embrace moral truth but deny its objectivity, and *relativists*, who deny its mind-independence. For the purposes of my argument, I will not

endeavor to prove any particular metaethical position. Rather, I will simply attempt to argue against positions that posit moral facts.

Subjectivism and Relativism

Nowadays, these positions are quite unpopular, and no wonder: they seem ill-posed to make sense of our moral talk, and they yield extremely counterintuitive consequences. Subjectivism says that moral statements are reports (not expressions) of a person's attitudes; so a subjectivist will construe a statement like "kicking puppies is wrong" as akin to a statement like "I (or whoever the speaker happens to be) disapprove of kicking puppies." On this simple subjectivist account, it will turn out that no two people actually mean the same thing when they make a moral judgment. But this in turn makes moral disagreement entirely inexplicable: if one person says that an action is right and another says it's wrong, the subjectivist will have to insist that there is no incompatibility between these claims (they are simply two different reports about two different people). Given such an implausible result, subjectivism seems untenable.

Relativism, which says that the truth of moral statements is relative to communities or cultures, also yields deeply unpalatable consequences. By way of elaboration, consider the Euthyphro Dilemma, which is sometimes posed against Divine Command Theorist.

Euthyphro Dilemma: Does God command certain actions because they're right? Or are they right because God commands them?

If the Divine Command Theorist wishes to persevere their theory, it seems they must opt for the latter option, but the point of the dilemma is that this is extremely odd: it implies that murder would have been acceptable if God had simply willed it. It is widely felt, however, that morality cannot be malleable in this way. And the point in the present context is that relativism implies exactly that - in fact, the relativist arguably has it worse than the divine command theorist, for whereas the latter can at least tie moral truth to the will of a wise, all-powerful, cosmic being, the former must say that morality is at the mercy of frail, ever-changing, finite custom.

Realism

In arguing against subjectivism and relativism, I attempted to show that their consequences drastically contravened common sense. Realism, however, *is* the common sense view of morality, so a more sophisticated rebuttal is required. To this end, I will summarize and defend what I take to be a very persuasive line against realism, viz. "the queerness argument" put forth by J.L. Mackie (Mackie, 1977).

The argument: The essence of Mackie's argument is that moral facts are objectionably strange, and so we shouldn't posit them. Of course, such facts might be admitted if they played an indispensable explanatory role in elucidating the way we think and talk, but they don't. There are clear and compelling accounts of how culture and biology shape our moral judgments, none

of which require the existence of moral truth. In fact, the anti-realist is arguably *better* situated than the realist to explain a stark feature of moral discourse: the prevalence of disagreement. Whereas the anti-realist can chalk up disputes to differences of sentiment or practice, the realist must explain why so many people, including the thoughtful and reflective among them, seem congenitally incapable of converging on the *truth*, even given ample time to jointly deliberate.

But what's so odd about moral facts? It's that they institute *categorical* imperatives, demands that are to be honored, regardless of one's desires or goals. By contrast, a *hypothetical* imperative tells you what you should do, given some specific desire. There is nothing especially strange about hypothetical imperatives; in essence, they simply describe the optimal means of achieving a goal, which is readily comprehensible in terms of empirical facts (e.g., one might say, "if you want to get there as soon as possible, you should take 2nd avenue", then elaborate by saying, "according to traffic data, 2nd avenue has shorter red lights and very few cars.") Categorical imperatives, however, cannot be understood in this way. It's as if they tell you how to achieve, not any particular person's preference, but the *world's* preference - which highlights the strangeness that Mackie and I find objectionable.

The 'companions in guilt' (CIG) objection: Some philosophers reject Mackie's argument on the grounds that it overreaches and implicates epistemic normativity as well. It's plausible, for example, that certain pieces of evidence constitute reasons for believing certain propositions - not hypothetical reasons, but categorical reasons, for they tell you what to believe *regardless of your desires*. Surely, if Mackie's reasoning is sound, then we ought to deny that such reasons exist. But that would be absurd, and so we should deny Mackie's argument instead.

My reply: I begin by disambiguating the term 'reason to believe', which can mean either:
(1) *a consideration that speaks in favor of a proposition's truth.*
(2) *a consideration that speaks in favor of adopting a belief, I.e., which shows why holding the belief is valuable.*

A reason in the first sense tells you that a proposition *p* is or is likely to be true, but it does not, as it were, tell you what to *do* with this information. It furnishes no prescriptions and does not designate any particular state of the world as better than any other; in particular, it does not single out the state of the world in which you believe that *p* as in any sense preferable to one in which you don't. Why is it called a 'reason *to believe*'? I suggest that's because, for virtually everyone in normal circumstances, the fact that a proposition is (or is likely to be) true is, at the same time, a reason to adopt belief in it. In other words, a reason in the first sense almost always functions as a reason in the second sense too. Strictly speaking, however, only the latter is a reason *to believe* (I.e., a reason to *do* something).

Now, in the CIG objection, what sense of 'reason to believe' is at stake? If the first, then I grant that such a reason does not depend on desire - it will continue to speak in favor of a proposition's truth regardless of anyone's goals or wants - but since such a reason issues no prescriptions, it is unproblematic for Mackie and me. But if the CIG involves reasons in the

second sense (which *do* issue prescriptions), then I deny that they are categorical. Whether it is valuable for someone to adopt a certain belief will depend on their preferences (for instance, the fact that a proposition is true will only count in favor of adopting belief in it *if* one prizes truth). Thus, I do not believe the CIG objection succeeds.

A Skeptical Regress

But even if Mackie's argument fails and moral realism is true after all, we still have no reason to trust any particular moral belief. For consider the following skeptical argument:

Premise 1: In order for a moral belief to be justified, it must either receive its justification from other beliefs or be self-justifying.

Premise 2: No moral belief is self-justifying.

Premise 3: In order for belief B to justify a moral belief, B must itself be a moral belief.

Now, consider some moral belief B1. By premise 2, B1 is not self-justifying, so by premise 1, if B1 is to be justified it must receive its justification from another belief B2, which by premise 3 must be a moral belief. But then, by the exact same logic, if B2 is to be justified it must receive its justification from another moral belief B3. And so on. But clearly this cannot go on forever, which means that B1 (which was arbitrary) is not justified. Consequently:

Conclusion: No moral belief is justified.

Elaboration and replies to objections: Premise 1 may be resisted by coherentists, who hold that justification consists in coherence with a wider belief set. However, I find a coherentist account of moral justification extremely implausible, for it seems one could construct a consistent ethical system featuring injunctives that most would consider abjectly *immoral* - would a realist admit that adherents of such a system are no less justified than everyone else? I doubt it, and thus I reject the objection.

Premise 3 follows from the Humean dictum that one cannot derive an 'ought' from an 'is'. To justify a moral belief (I.e., a belief about how one *ought* to act), one cannot appeal to how things *are*. Rather, one will have to cite another belief about how things ought to be - which is to say, another moral belief.

Against premise 2, it may be insisted that some moral beliefs are genuinely self-justifying. However, it is difficult to see what 'self-justification' could amount to in this context. Are moral propositions 'analytic', I.e., true in virtue of their meanings (like the self-justifying proposition, "all bachelors are unmarried")? Are they self-evident (like the self-justifying proposition "2+2=4")? No, for as mentioned before, moral propositions are subject to vast and persistent disagreement - which would be baffling if merely comprehending them was sufficient to recognize their truth.

It might be urged that moral self-justification is akin to perceptual self-justification: just as people need no reason to trust their senses, they are in the same position with respect to (at least some of) their moral judgments. But I believe the analogy is inapt, for whereas doubts about sense-perception are almost invariably fanciful (a la Descartes' "evil demon") and essentially impossible to sustain in practice, doubts about morality stem from grounded and realistic philosophical and scientific premises. This is not to say that moral skepticism is necessarily true, but merely that its relative plausibility (in contrast to the utter *implausibility* of external world skepticism) renders moral beliefs in need of reasons in a way that perceptual beliefs are not.

Conclusion

Thus, there is no reason to trust moral intuition, for to do so presupposes that (1) there are moral facts, and (2) some of our moral beliefs are justified. As I've attempted to show, however, both (1) and (2) are false.

Works Cited

Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Penguin Books.

A Weapon of Mass Distraction: Misinformation in the Russian-Ukrainian Conflict

By Jerome Ong

ABSTRACT

Tackling the misinformation crisis is a perennial challenge that has stumped policymakers for decades. With the advent of social media, this issue has become even more pertinent for discussion. Extensive research has been conducted to broadly assess the misinformation crisis or recommend specific policies in unique circumstances. Yet, there has been little investigation into specific case studies of misinformation to determine the future of misinformation regulation. As such, this research paper will predominantly use the context of the Russian-Ukrainian war to discuss the scope and magnitude of the misinformation crisis. Subsequently, it discusses the strengths and limitations of current misinformation regulation before providing a general direction for future regulation, drawing on lessons from the Russian-Ukrainian War.

INTRODUCTION

Digital misinformation has emerged as one of the most destabilising forces in the last decade, no longer confined to slow-moving rumours or hearsay.¹ Through a hyper-connected digital ecosystem, misinformation now fuels the agenda of bad actors, sows the seed of fear among the populace, and even threatens to destroy the social fabric of society as we know it today. It is uniquely characterised by its speed, reach, and integration with an indispensable part of our lives – social media.² As such, misinformation has been transformed from a peripheral risk into a systemic threat that needs to be addressed immediately. This is particularly evident in the 2022 Russian-Ukrainian war, which has even been described as the “first social media war” in history.³ While misinformation exists on both sides of the conflict, this research will predominantly analyse pro-Kremlin misinformation, given evidence suggesting that misinformation has been used as a strategic weapon to influence public sentiment and project military power.⁴

¹ Dean Jackson and Jon Bateman, “Countering Disinformation Effectively: An Evidence-Based Policy Guide | Carnegie Endowment for International Peace,” accessed October 21, 2025, <https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en>.

² Evelyn Douek, “Content Moderation as Systems Thinking,” *Harvard Law Review* 136, no. 2 (2022).

³ Dan Ciuriak, “The Role of Social Media in Russia’s War on Ukraine,” SSRN Scholarly Paper no. 4078863 (Social Science Research Network, April 8, 2022), <https://doi.org/10.2139/ssrn.4078863>.

⁴ Aaron Erlich and Calvin Garner, “Is Pro-Kremlin Disinformation Effective? Evidence from Ukraine,” *The International Journal of Press/Politics* 28, no. 1 (2023): 5–28, <https://doi.org/10.1177/19401612211045221>.

To understand the importance of misinformation regulation, Section 1 will first define the scope of misinformation and clarify the distinctions among misinformation, disinformation, and misleading information. Subsequently, it will lay out the magnitude of harm caused by misinformation, namely through the creation of inaccurate narratives and the formation of echo chambers. By further examining the international scale and deep-rooted tensions surrounding the Russian-Ukrainian war, this section will also stress why regulation has become a pressing necessity.

Building on this foundation, Section 2 will analyse the two main types of misinformation in the Russian-Ukrainian war: military- and politically-related. While this research has already established that misinformation is systematically amplified by social media algorithms, engineered to provoke an emotional response, and capable of influencing millions, this section seeks to evaluate the various methods and media used to spread misinformation in a contemporary geopolitical landscape. Lastly, Section 2 will introduce society's current response to misinformation and ultimately demonstrate that the absence of social media regulation will inevitably lead to its spread.

Section 3 directly examines the contentious debate over social media regulation, evaluating the strengths and limitations of self-regulation and state regulation. Through global case studies and examples from the Russian-Ukrainian war, Section 3 will highlight the trade-off between freedom of speech and efforts to reduce harmful misinformation. This sheds light on the complex and dynamic challenge of misinformation regulation whilst considering the various motivations of stakeholders.

Finally, Section 4 will consolidate the findings from the first three sections and holistically analyse the nuances of the misinformation crisis in the Russian-Ukrainian war. This paper will not only evaluate current efforts to regulate misinformation but also discuss a sustainable path forward. This paper aims to strengthen global resilience to misinformation, thereby protecting the social fabric that underpins both democratic governance and civic stability.

SECTION 1: THE MISINFORMATION CRISIS

To understand the importance of misinformation regulation, Section 1 will outline what constitutes misinformation, before further contextualising it in relation to the Russian-Ukrainian war. This section aims to fulfil these objectives by elucidating the detrimental impacts and immense scale of the misinformation crisis. Coupled with the complicating factors of the Russian-Ukrainian war, the need for social media regulation will become apparent.

Section 1.1: The scope of misinformation

The definition of misinformation varies across sources. Caldwell University defines it as “any incorrect information that is shared without an intent to deceive.”⁵ At the same time, the

⁵ Christina Getaz, “Research Guides: EN 101: The Process of Writing: Misinformation & Disinformation,” accessed October 12, 2025, <https://libguides.caldwell.edu/EN101/misinformation-disinformation>.

Cambridge dictionary further extends this definition to include any situation in which “people are misinformed.”⁶ Princeton University even refers to misinformation as “a catch-all” for all types of false or inaccurate information, regardless of whether the act of referring to or sharing it was intentionally misleading.⁷

Some sources distinguish between misinformation and disinformation, with intent being the main difference. As stated earlier, Caldwell University defines misinformation as the spread of misleading information without deceptive intent. Still, it defines disinformation as “false or misleading information shared with the intent to mislead, confuse, or deceive people.”⁸ While intent to mislead accurately distinguishes between misinformation and disinformation, discerning intention in the real world is complicated. For instance, a journal discussing consumer demand might report a ten percent increase in spending over the last five years. Although only one digit off the actual figure, stating that there is “a hundred percent increase in spending” will carry a markedly different connotation. Yet it is difficult to determine whether a scientific journal accidentally added an extra zero to the figure or deliberately overstated the growth in consumer demand to make its point.

To facilitate discussion, this paper defines misinformation as false or inaccurate information, regardless of intent. Given that the primary focus of this paper is the mechanics of regulating misleading information on social media platforms, an exploration of intent would be unproductive and require separate research. Therefore, by addressing both unintentional and malicious fake news, this streamlines the discussion of information classification (whether it is traditionally considered misinformation or disinformation), allowing for more in-depth evaluation of how to regulate misleading information.

Section 1.2: The magnitude of harm caused by misinformation

The danger of misinformation lies in its ability to remain indistinguishable from factual information. Given that it is premised on misguided or misleading information, it often relies on emotional manipulation to erode our ability to gather accurate and reliable information. Its harms manifest in two dominant ways, namely the clouding of objective facts and the creation of echo chambers.

Firstly, misinformation clouds objective facts by creating inaccurate narratives. When social media posts feel like factual information, readers tend to respond strongly. The problem arises when emotional responses to the information outweigh its veracity.⁹ This allows readers to

⁶ Cambridge University, “MISINFORMATION | English Meaning - Cambridge Dictionary,” accessed July 27, 2025, <https://dictionary.cambridge.org/dictionary/english/misinformation>.

⁷ Princeton University, “Misinformation, Disinformation & Malinformation: A Guide - Princeton Public Library,” accessed July 27, 2025, <https://princetonlibrary.org/guides/misinformation-disinformation-malinformation-a-guide/>.

⁸ Getaz, “Research Guides.”

⁹ Magdalene Karalis, “Russia-Ukraine War Through the Eyes of Social Media | GJIA,” *Georgetown Journal of International Affairs*, February 2, 2024, <https://gjia.georgetown.edu/2024/02/02/russia-ukraine-through-the-eyes-of-social-media/>.

form emotional connections to misinformation, leading them to believe a false narrative even before fact-checking. Even in the absence of emotional connection, misinformation also benefits from the rapid diffusion of content through social media.¹⁰ It is significantly easier to produce misinformation than factual information, because factual information requires testing and verification, whereas misinformation relies on emotional rhetoric to appeal to the audience. This results in a greater volume of misinformation than factual information, leading to the oversaturation of misinformation on social media. This information overload leads the audience to deprioritise the veracity of claims, focusing instead on features that improve audience retention, such as emotive language. This clouds objective facts and makes the audience more susceptible to believing misinformation.

Secondly, misinformation is problematic because it creates echo chambers. Social media algorithms work by recommending to users what they most likely want to watch. This means that the platform prioritises content that garners user engagement – regardless of its accuracy. Consequently, users are drawn into information silos, where they are exposed only to content that already reinforces their preexisting beliefs, even when those beliefs were borne of misinformation.¹¹ In the context of the Russian-Ukrainian war, this appears to be a curious user seeking more information about the conflict by watching war-related videos on social media. At some point, an assertive claim that the Russians were justified in starting the war attracts the user's attention. This causes him to spend more time reading that piece of content, leading the algorithm to conclude that this type of content increases engagement for this user. Over time, the content fed to this user becomes increasingly one-sided, disproportionately glorifying the role of the Russians. This example illustrates how social media's recommendation algorithms tend to amplify engaging content, even when it is factually incorrect.¹² At the same time, the formation of echo chambers surrounds users with a community that shares and legitimises the same view, making it harder for them to recognise that some content stems from misinformation.

Apart from the apparent problems propagated by misinformation, identifying whether a particular piece of information constitutes misinformation is also frustratingly obtrusive. One possible reason for this could be the blurry line between fact and opinion. Just as it is difficult to discern the intent behind spreading misinformation, it can be difficult to distinguish between facts and opinions. When an individual wants to present an objective fact on social media, there

¹⁰ Joshua A. Tucker et al., "From Liberation to Turmoil: Social Media And Democracy," *Journal of Democracy* 28, no. 4 (2017): 46–59, <https://doi.org/10.1353/jod.2017.0064>.

¹¹ Megan Kurten, "Why We Fight for Fractured Truths – How Misinformation Fuels Political Violence in Democracies," *MediaWell*, January 9, 2025, <https://mediawell.ssrc.org/research-reviews/why-we-fight-for-fractured-truths-how-misinformation-fuels-political-violence-in-democracies/>.

¹² Clothilde Goujard, "YouTube's Algorithm Pushes Hateful Content and Misinformation: Report," *Politico*, July 7, 2021, <https://www.politico.eu/article/mozilla-firefox-report-youtube-algorithm-pushes-hateful-content-misinformation/>.

are still countless subjective decisions to make, such as the language used to demonstrate the point, the description of the fact, or even the balance between textual and visual elements.¹³ There is a non-zero chance that the resulting content is misinformation, especially since these subjective elements can be interpreted differently based on a viewer's worldview or lived experiences, thereby leading to inaccurate interpretations of the information.¹⁴

Another reason for this difficulty stems from the difficulty of discerning fact from fabrication. By definition, short-form content on social media can only provide abridged explanations or descriptions. For instance, this could look like thirty-second videos with three to four lines of speech explaining why climate change is a hoax, as well as posts featuring aesthetic infographics on healthy eating that claim to be based on "scientific research." Although these simplified narratives have innocuous intentions, they can cause significant problems when used to substantiate broader, nuanced arguments. With the information posted being highly generalised, it is hard to identify when the content is based on semblances of the truth or pure fabrication.¹⁵

Section 1.3: The scale of harm caused by misinformation

Beyond the immense magnitude of harm it can cause, the speed and scale of misinformation's spread underscore the need for social media regulation. According to Pew Research Center research, social media now plays a central role in news consumption. In 2024, fifty-four percent of American adults get their news from social media, underscoring how digital sources have become an essential part of individuals' news diets.¹⁶

This is especially relevant when discussing misinformation stemming from the Russian-Ukrainian conflict, specifically. With millions of lives displaced, global supply chain disruptions, and shifting global power dynamics, the Russian-Ukrainian conflict is one of the highest-stakes conflicts of recent history. Given that nationalistic sentiments are being fuelled by strong emotions such as pride and fear, online discourse around this conflict becomes emotionally charged. This subconsciously influences individuals to take sides in the conflict, leading to a high degree of politicisation on social media platforms.¹⁷

¹³ Stefania Farace et al., "Standing Out While Fitting In: Visual Design of Text Overlays in Social Media Communication," *Journal of Marketing*, SAGE Publications Inc, February 12, 2025, 00222429251322773, <https://doi.org/10.1177/00222429251322773>.

¹⁴ Shiqi Yi et al., "Is Information Evaluated Subjectively? Social Media Has Changed the Way Users Search for Medical Information," *DIGITAL HEALTH* 10 (September 2024): 20552076241259039, <https://doi.org/10.1177/20552076241259039>.

¹⁵ Sierra L. Lawson and Steven W. Ramey, "Sourcing Stereotypes: Constructing and Challenging Simplified Knowledge," *Culture and Religion* 19, no. 4 (2018): 416–34, <https://doi.org/10.1080/14755610.2018.1505759>.

¹⁶ Pew Research Center, "Social Media and News Fact Sheet," *Pew Research Center*, September 17, 2024, <https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/>.

¹⁷ Gianella Pena Barbary, *The Role of Communication in 21st Century Nationalism: A Tool of Influence and Manipulation*, April 30, 2025,

Even though many other conflicts in recent history have suffered heavy casualties (such as the Yemen or Ethiopian crisis), none have had media coverage on par with that of the Russian-Ukrainian conflict. As such, the Russian-Ukrainian war has been described as the first “social media war” in history.¹⁸ Every development, atrocity, and counter-narrative is broadcast worldwide through social media, with posts and comments providing additional updates on the war. The virality of such content produced by Ukrainians, Russians, and their respective sympathisers makes the situation even more volatile, demonstrating the global reach of the Russian-Ukrainian conflict.

To the global community, the Russian-Ukrainian war is more than a regional conflict. It has the potential to escalate to an international conflagration with superpowers taking sides.¹⁹ However, one must also recognise that the tensions between both countries have existed long before Russia’s direct invasion of Ukraine in 2022. This means that there has been, and will continue to be, exceptional scrutiny of the Russian-Ukrainian conflict, thereby increasing demand for information and the reach of misinformation surrounding this issue. Therefore, now more than ever, it is crucial to keep misinformation under control to prevent its harmful effects from affecting a significant portion of society that relies on social media for news.

However, this will be challenging given recent discoveries regarding the spread of misinformation. A study conducted by MIT reveals how false news spreads much faster than real news by a substantial margin. Across all categories of information, falsehoods diffused significantly faster, deeper, and farther than factual information.²⁰ To rationalise these findings, this paper provides two possible explanations.

On one hand, the emotional charge of misinformation may drive audience engagement. Unlike factual information, which should always aim to make a rational point, misinformation leverages the power of strategic narratives to appeal to viewers’ emotions.²¹ Effectively, instead of convincing users with logic and facts, misinformation taps into emotions to blur the line

<https://identityhunters.org/2025/03/30/the-role-of-communication-in-21st-century-nationalism-a-to-ol-of-influence-and-manipulation>.

¹⁸ Dan Ciuriak, “Social Media Warfare Is Being Invented in Ukraine,” *Centre for International Governance Innovation*, n.d., accessed April 21, 2025, <https://www.cigionline.org/articles/social-media-warfare-is-being-invented-in-ukraine/>.

¹⁹ Malaka Gharib, “Not Every War Gets the Same Coverage as Russia’s Invasion — and That Has Consequences,” *Ukraine Invasion — Explained*, *NPR*, March 4, 2022, <https://www.npr.org/sections/goatsandsoda/2022/03/04/1084230259/not-every-war-gets-the-same-coverage-as-russias-invasion-and-that-has-consequenc>.

²⁰ Peter Dizikes, “Study: On Twitter, False News Travels Faster than True Stories,” *MIT News*, March 8, 2018, <https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308>.

²¹ Irina Khaldarova and Mervi Pantti, “Fake News: The Narrative Battle over the Ukrainian Conflict,” *Journalism Practice* 10, no. 7 (2016): 891–901, <https://doi.org/10.1080/17512786.2016.1163237>.

between reality and fabrication, thereby facilitating easier message retention. Thus, even if viewers doubt the accuracy of such content, the emotional rhetoric leaves a lasting impression. This looks like examples such as the 1998 “The Lancet” paper written by Andrew Wakefield. The paper fraudulently reported that the measles, mumps, and rubella (MMR) vaccine contributed to a higher chance of autism in children. While the fraud was quickly brought to light, the shocking nature of the claim left a lasting impression on countless individuals worldwide. Even to this day, society is dealing with its social impact, as many continue to wonder whether this statement holds some truth.²²

Alternatively, it can also be argued that platforms themselves facilitate the spread of misinformation. This is because social media companies are inherently profit-driven. Given that these companies can only achieve their goals by increasing users' time spent on social media platforms, they would naturally prioritise user engagement. Yet this creates a perverse incentive to encourage users to share content indiscriminately, regardless of quality or accuracy, because misinformation can be more engaging for the reasons stated earlier.²³ Therefore, this paper posits that the spread of misinformation may be encouraged on a systemic level as well.

SECTION 2: QUANTIFYING THE PROBLEM

Building on the earlier characterisation of misinformation, Section 2 explores the different types of misinformation in the status quo. It will also discuss the methods and media used to spread misinformation, along with other difficulties in regulating the status quo. This will effectively illustrate the problem of misinformation, providing context on its various forms and the public’s response to it in the status quo.

Section 2.1: Types of misinformation

Misinformation can vary in type and nature. From inaccurate data in viral videos to doctored images on social media posts, there are countless ways in which misinformation has emerged in the Russian-Ukrainian war. Nevertheless, for the sake of discussion, this paper has identified two dominant categories of misinformation: Military-related misinformation and political misinformation. These categories are not merely academic; they encapsulate real events.²⁴

On one hand, military-related misinformation deserves particular attention in this conflict. Given the direct military confrontation taking place in Ukraine, it is only logical that misinformation regarding the war efforts starts to spread through social media. Without being physically present in Ukraine, the international community lacks firsthand knowledge about the conflict. Instead, they are forced to piece together scraps of information from social media to

²² Laura Eggertson, “Lancet Retracts 12-Year-Old Article Linking Autism to MMR Vaccines,” *CMAJ: Canadian Medical Association Journal* 182, no. 4 (2010): E199, <https://doi.org/10.1503/cmaj.109-3179>.

²³ Gizem Ceylan, “How Social Media Rewards Misinformation,” *Yale Insights*, March 31, 2023, <https://insights.som.yale.edu/insights/how-social-media-rewards-misinformation>.

²⁴ Erlich and Garner, “Is Pro-Kremlin Disinformation Effective?”

understand the conflict's developments better.²⁵ This scarcity of information has led users to misrepresent or outright fabricate information to contribute to the conversation about the Russian-Ukrainian war. Studies have also revealed how Russia actively weaponizes social media as part of its pro-Kremlin disinformation strategy, exaggerating its defeats and fabricating war crimes, all to mislead the international community. Russia portrays itself as a military liberator, while discrediting NATO and other Western institutions as imperialist aggressors. Unfortunately, as Greenhill and Oppenheim suggest, in the absence of direct evidence, individuals are inclined to accept disinformation rather than no information at all.²⁶ This leads to the subconscious acceptance of disinformation as truth, which can subsequently cause users to propagate misinformation themselves, now convinced they know the truth.

Political misinformation is another category for consideration. Beyond misinformation directly targeting the war efforts in Ukraine, there is also misinformation regarding the context of the conflict. Russia spreads narratives of its country being the heroic victor of World War II, and Ukraine betraying its historical roots by aligning with the West. This narrative becomes even more believable when considering that some of the accusations against Ukraine have proven true. The Azov Brigade, for example, was accused of having Nazi connections given its contentious use of Nazi-related symbols, and members of its unit committing human rights violations.²⁷ This has led to the rampant spread of misinformation, whether intentionally by the Russian government or unintentionally by members of the public, that the rest of Ukraine is similarly fascist or has Nazi connections. Regardless of intent, this misinformation has tainted Ukrainian leaders with suspicious historical connections.²⁸ The scandalous nature of such claims has further spread misinformation. Even though a claim may be supported by dubious sources or fabricated evidence, the initial outrage still garners traction on social media, depicting Ukraine as an illegitimate and morally grey country.

Both political and military misinformation pose a similar threat to society. They tend to promote pro-Kremlin narratives, conveying the consistent narrative that Russia will eventually “liberate” Ukraine, whether it is by exaggerating the overwhelming military resources of the Russians, or tainting the image of Ukrainian leaders through unsavoury historical connections.²⁹ Thus, these two types of misinformation will be referenced throughout the rest of the research.

Section 2.2: Methods of misinformation

²⁵ Erlich and Garner, “Is Pro-Kremlin Disinformation Effective?”

²⁶ Kelly Greenhill and Ben Oppenheim, “Rumor Has It: The Adoption of Unverified Information in Conflict Zones,” *International Studies Quarterly* 61, no. 3 (2017): 660–76, <https://doi.org/10.1093/isq/sqx015>.

²⁷ Al Jazeera, “Profile: Who Are Ukraine’s Far-Right Azov Regiment?,” Al Jazeera, accessed October 5, 2025, <https://www.aljazeera.com/news/2022/3/1/who-are-the-azov-regiment>.

²⁸ Erlich and Garner, “Is Pro-Kremlin Disinformation Effective?”

²⁹ Erlich and Garner, “Is Pro-Kremlin Disinformation Effective?”

Even on social media, misinformation can be spread through various methods. Among the most common are repurposed photos and videos, fake eyewitness accounts, and internet trolls. Each of these methods contributes to misinformation on social media in its own way.³⁰

Firstly, recontextualised images from unrelated sources are used to support false narratives, misdirecting viewers and leading them to form incorrect conclusions. For instance, a war photo in 2014 was falsely circulated as proof that Ukrainian soldiers injured a twelve-year-old boy in Sloviansk, thereby acting as evidence of Ukrainian aggression in Sloviansk. However, it was later revealed that the photo was a 2013 photo from the Syrian War, with no connection to the supposed Ukrainian aggression.³¹ This method of misinformation is particularly effective due to the well-documented Picture-Superiority Effect (PSE). First discovered by Paivio and Csapo in 1973, PSE is a cognitive bias in which images are more likely to be internalised than text on a screen.³² Even for bite-sized information where both text and image are easily remembered in immediate recall, images were found to be significantly more memorable in delayed recall.³³ At the same time, unlike other methods of misinformation, which present facts and data to viewers, images serve as visual stimuli that influence viewers' interpretation of the situation. Using contextual clues from the surrounding text, viewers form connections between text and image, under the tacit assumption that images must be relevant to the accompanying text.³⁴ However, when images are recontextualised, their context is misrepresented. Viewers end up with an inaccurate interpretation, thereby propagating misinformation.

Additionally, fake eyewitness accounts lend a false sense of legitimacy, further propagating misinformation.³⁵ In 2014, a self-proclaimed pro-Russian emergency doctor, Igor Rozovski, reported on Facebook that Ukrainian nationalists did not allow him to help other injured pro-Russian demonstrators after a fire at the Trade Unions House. The story was reposted thousands of times and even translated into different languages before it was realised that

³⁰ Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (Yale University Press, 2018), <https://doi.org/10.12987/9780300235029>.

³¹ Gillespie, *Custodians of the Internet*.

³² Abd Allah Aljalabneh, "Visual Media Literacy: Educational Strategies to Combat Image and Video Disinformation on Social Media," *Frontiers in Communication* 9 (November 2024), <https://doi.org/10.3389/fcomm.2024.1490798>.

³³ Terry L. Childers and Michael J. Houston, "Conditions for a Picture-Superiority Effect on Consumer Memory," *Journal of Consumer Research* 11, no. 2 (1984): 643–54, <https://doi.org/10.1086/209001>.

³⁴ Eryn J. Newman and Norbert Schwarz, "Misinformed by Images: How Images Influence Perceptions of Truth and What Can Be Done about It," *Current Opinion in Psychology* 56 (April 2024): 101778, <https://doi.org/10.1016/j.copsyc.2023.101778>.

³⁵ Gillespie, *Custodians of the Internet*.

Rozovskiyy was not an emergency doctor but a dentist. Additionally, his Facebook account was created only immediately before he published that story and was deleted soon after that truth was revealed.³⁶ While the truth was eventually revealed, it was not before thousands believed the fabricated incident. What makes this method of misinformation especially problematic is the lack of readily available corroborative evidence. To correct misinformation, the correction needs to be specific enough for the public to understand which aspect of the claim is being refuted.³⁷ However, in the case of the Russian-Ukrainian War, verifying or refuting statements is challenging due to the conflict's isolated nature. Therefore, so long as the information seems logically congruent, individuals tend to believe the claim, regardless of its actual veracity.

Nevertheless, the most pervasive source of misinformation remains internet trolls. Internet trolls provoke social discord by expressing disagreement towards the subject matter through a combination of insults, jokes, and unhelpful comments.³⁸ Such trolls have various motivations for spreading misinformation, ranging from being excessively opinionated to being a state-sponsored agent or simply wanting entertainment. Russia has repeatedly denied the existence of state-sponsored internet trolls, but the discovery of a “troll factory” situated near St. Petersburg suggested otherwise. Regardless of whether the Russian government indeed paid the trolls, their disruptive effect is undeniable. The Guardian reported coordinated attacks on its Russia and Ukraine articles, facing up to forty thousand troll comments per day.³⁹ Internet trolls are particularly effective as they rely on quantity, rather than the quality of information. By flooding comment sections, mass-reporting posts, and asserting unsubstantiated claims, internet trolls drown out legitimate, factual information, resulting in the rapid propagation of misinformation.⁴⁰

Section 2.3: Media of misinformation

At the same time, this paper acknowledges that not all social media platforms are identical: Some platforms, like Telegram, are vastly different from others like Instagram or TikTok. Understanding these differences will be crucial, as they have consequential effects on social media regulation and inform policy decisions on how best to regulate misinformation across platforms.

One medium for misinformation would be social messaging applications, such as Telegram. The platform has been widely used during the Russian-Ukrainian war to

³⁶ Gillespie, *Custodians of the Internet*.

³⁷ Toby Pike and Ullrich K. H. Ecker, “Effective Correction of Misinformation,” *Current Opinion in Psychology* 54 (December 2023): 101712, <https://doi.org/10.1016/j.copsyc.2023.101712>.

³⁸ Evita March and Jessica Marrington, “A Qualitative Analysis of Internet Trolling,” *Cyberpsychology, Behavior, and Social Networking* 22, no. 3 (2019): 192–97, <https://doi.org/10.1089/cyber.2018.0210>.

³⁹ Gillespie, *Custodians of the Internet*.

⁴⁰ Flemming Splidsboel Hansen, “Trolls in Your Feed - Russian Disinformation,” accessed October 5, 2025, <https://www.diis.dk/en/trolls-in-your-feed-russian-disinformation>.

disseminate information and boost morale. Over the last few years, Ukrainian soldiers have posted high-intensity fighting videos and drone footage showing attacks on the Russian troops in Telegram chats, helping to raise soldiers' spirits and appeal to the moral obligation of international leaders.⁴¹ On the other hand, the Russian government has flooded the platform with over three hundred thousand inauthentic comments under the guise of seeking peace and cooperation.⁴² These comments promote pro-Russian ideology and anti-Ukrainian narratives targeted at Ukraine's occupied territories, fuelling the government's efforts to control the narrative by framing the Russian-Ukrainian war as a "special military operation."

Regardless of intention, the decentralised spread of information via text messages, pictures, and videos creates a high likelihood of misinformation on Telegram. This is primarily due to end-to-end encryption, which ensures that messages are unreadable by third parties, including fact-checkers who would otherwise be able to track the content being shared.⁴³ Nevertheless, this medium for misinformation has the lowest impact. This is because the spread of misinformation through this medium is reliant on the viewer's interest in the subject. Unlike algorithm-based platforms, Telegram requires users to actively join channels or group chats to receive information.⁴⁴ Therefore, if an informed individual recognises the risk of misinformation from a specific source, they will consciously choose not to join the channel, thereby breaking the chain of misinformation.

Another medium for misinformation would be media-sharing platforms like Instagram and TikTok. These platforms are highly visual and prioritise user engagement through their algorithm-driven recommendation system. Additionally, studies have shown that user engagement is prolonged by emotionally charged, shocking, or even polarising elements.⁴⁵ Thus, it is unsurprising that there have been numerous instances in which these platforms have been used to post aestheticized war footage with misleading captions and edited clips from interviews with government officials. For example, a study by BBC Verify has exposed many videos on TikTok featuring high-ranking government officials, including President Volodymyr Zelensky, being unfairly portrayed as driven by profit and detached from the suffering of the Ukrainians in

⁴¹ Karalis, "Russia-Ukraine War Through the Eyes of Social Media | GJIA."

⁴² Atlantic Council, "Digital Occupation: Inside Russia's Telegram Battle in Ukraine - Atlantic Council," July 16, 2025, <https://www.atlanticcouncil.org/event/digital-occupation-inside-russias-telegram-battle-in-ukraine/>.

⁴³ BBC, "Ukraine Bans Telegram Use on State-Issued Devices," September 21, 2024, <https://www.bbc.com/news/articles/c78dwepw95do>.

⁴⁴ Telegram, Telegram, "Telegram Channels," accessed October 6, 2025, <https://telegram.org/tour/channels>.

⁴⁵ Zicheng Cheng and Yanlin Li, "Like, Comment, and Share on TikTok: Exploring the Effect of Sentiment and Second-Person View on the User Engagement with TikTok News Videos," *Social Science Computer Review* 42, no. 1 (2024): 201–23, <https://doi.org/10.1177/08944393231178603>.

the war.⁴⁶ The spread of misinformation through this medium is especially detrimental, as the algorithm-based recommendation system is highly likely to continue exposing users to misleading political content, even without seeking it out, so long as they have watched similar content on the Russian-Ukrainian conflict.

Another medium for misinformation is discussion forums, including Twitter and Reddit. These platforms offer a conversation-like format for individuals to share opinions, view links, and join communities based on various topics.⁴⁷ In the context of the Russian-Ukrainian conflict, this looks like users sharing absurd claims such as the Russian-Ukrainian War being fake.⁴⁸ Unlike the other two mediums, misinformation is not directly forwarded to users or recommended by the algorithm. Instead, users actively choose the topic of discussion and interact with other users on the platform. Unfortunately, these forums prioritise group validation and reinforcement over information accuracy through their upvote system. Given the earlier characterisation of how emotional and political charges can overshadow the veracity of information, these discussion forums turn dubious claims into widely accepted truths if enough community members upvote them.⁴⁹

Section 2.4: The current response to misinformation

Given the variety of misinformation types, methods, and media, how does society currently respond to the misinformation crisis?

Firstly, bad actors increasingly weaponize social media to spread misinformation. These actors recognise how social media enables them to create spectacle rapidly, while drawing emphasis away from the lack of veracity of some of these claims.⁵⁰ As a result, misinformation has become more pervasive than ever. These actors continue to blur the boundary between truth and fiction, creating a post-truth environment where emotional appeal dominates social media.

As for regular social media users, a growing proportion of society is cognisant of the issue of misinformation, though to varying degrees. For most people, this knowledge comes in the form of a mindset shift. Now, when an individual comes across a dubious image on social media, their first thought would be: “Is it AI?” Such instinctive responses reveal how, as a society, we are beginning to acknowledge the pervasiveness of misinformation and its impact on our day-to-day lives. Even in the context of the Russian-Ukrainian war, users are aware of the

⁴⁶ BBC, “*Ukraine War: How TikTok Fakes Pushed Russian Lies to Millions*,” December 15, 2023, <https://www.bbc.com/news/world-europe-67687449>.

⁴⁷ Katie Elson Anderson, “Ask Me Anything: What Is Reddit?,” *Library Hi Tech News* 32, no. 5 (2015): 8–11, <https://doi.org/10.1108/LHTN-03-2015-0018>.

⁴⁸ Alistair Coleman and Shayan Sardarizadeh, *Ukraine War: Viral Conspiracy Theories Falsely Claim the War Is Fake*, February 27, 2023, <https://www.bbc.com/news/world-europe-64789737>.

⁴⁹ Holly Haney, “Systems of Upvoting and Their Implications (Term Paper),” *Medium*, December 16, 2015, <https://hollyhaney.medium.com/systems-of-upvoting-and-their-implications-term-paper-1954a004e916>.

⁵⁰ Ceylan, “How Social Media Rewards Misinformation.”

pro-Kremlin disinformation campaign and remain largely skeptical of information from Russian sources.⁵¹

A smaller proportion of society goes even further. Given how this is not the first wave of misinformation, communities are already beginning to tackle misinformation through crowdsourced projects. For instance, StopFake was launched in 2014 in Kyiv to combat misinformation across various sources, including social media.⁵² Although it started as a small grassroots initiative to debunk, it has since been joined by numerous professionals, growing into a vibrant community and symbolising a new wave of citizen accountability in digital spaces. Although efforts thus far have been largely uncoordinated, the mere presence of such crowdsourced projects indicates an increasing desire to weed out misinformation.⁵³

At the same time, more vulnerable members of society continue to fall prey to misinformation. Based on the research conducted by Erlich and Garner, individuals who are less politically astute, identify with Russian ethnicity and language, and are demographically older tend to be more susceptible to misinformation. Parallels can be drawn to the status quo, with youths forwarding misleading reels on Instagram, or elders sending chain letters in their family WhatsApp chats.⁵⁴

In all cases, it is evident that some form of intervention is necessary, whether by the government or by social media companies. When left to their own devices, social media users often fall victim to misinformation. Even within the subset of users that recognise the importance of tackling misinformation, crowdsource projects are currently uncoordinated and ineffective.

SECTION 3: REGULATION IN THE STATUS QUO

The question of social media regulation has sparked immense discussion and controversy. Striking the right balance between protecting freedom of speech and curbing harmful misinformation requires both platform self-regulation and government-led regulation. Although both approaches have yielded some results, neither is without shortcomings. Section 3 will explore the state of regulation in these two areas, drawing on case studies and regulatory policies from around the world.

Section 3.1: Self-regulation

Social media companies are keenly aware of the misinformation crisis that plagues their platforms. As a result, companies engage in self-regulation, voluntarily identifying and removing misinformation without direct state intervention. Self-regulation often involves internal rules and regulations adopted by companies on their own accord. For example, this could include protocols

⁵¹ Khaldarova and Pantti, “Fake News.”

⁵² Gillespie, *Custodians of the Internet*.

⁵³ Gillespie, *Custodians of the Internet*.

⁵⁴ Erlich and Garner, “Is Pro-Kremlin Disinformation Effective?”

for scrutinizing advertisement placements, services, and fact-checking mechanisms.⁵⁵ At the same time, self-regulation is often guided by international frameworks and codes of conduct. In the EU, social media platforms such as YouTube and TikTok have committed to more accountable content moderation in line with the EU Code of Practice on Disinformation and the EU Code of Conduct on Countering Illegal Hate Speech Online. In 2022, the broader environment evolved with the Digital Services Act (DSA), imposing legally binding obligations on social media companies to combat the spread of harmful content.⁵⁶ This policy builds on years of past self-regulatory efforts by social media companies, streamlining such policies on an international scale.

Social media companies have an active incentive to self-regulate for various reasons. Some platforms conduct content moderation out of a sense of moral responsibility.⁵⁷ For other social media companies like YouTube and Twitter, establishing free-speech norms is crucial to maintaining their identities. As such, they have established community guidelines for misinformation aligned with the company's values and goals.⁵⁸ Even for social media companies that may not have a strong orientation towards managing content quality, content moderation will be necessary to prevent the backlash that can result from allowing misinformation to spread rampantly across the platform.⁵⁹ More compelling, however, are the economic incentives. Although corporate responsibility exists, the main reason most companies continue to regulate misinformation is the risk of losing potential profits. This is because allowing misinformation to propagate on the platform can lead to the alienation of users and advertisers, especially given the highly politicised and divisive nature of the Russian-Ukrainian war. This can lead to direct revenue losses when users leave the platform.⁶⁰ Therefore, to retain user engagement and protect the company's profitability, implementing an effective content moderation system is crucial.

There are many examples of self-regulation by social media companies. For instance, when companies identify misinformation, most social media algorithms actively reduce its visibility by not recommending it to users, thereby limiting its spread and disincentivising users

⁵⁵ Ekin Balkan and Sinan Ülgen, *THE ROLE OF SELF REGULATION AND REGULATION IN COMBATING DISINFORMATION*, n.d., https://edam.org.tr/uploads/yukleme_resim/Report-The-Role-of-Self-Regulation-and-Regulation-in-Combating.pdf.

⁵⁶ Balkan and Ülgen, *THE ROLE OF SELF REGULATION AND REGULATION IN COMBATING DISINFORMATION*.

⁵⁷ Jeffrey Howard, "The Ethics of Social Media: Why Content Moderation Is a Moral Duty," *Journal of Practical Ethics* 11, no. 2 (2024), <https://doi.org/10.3998/jpe.6195>.

⁵⁸ Kate Klonick, "The New Governors: The People, Rules, and Processes Governing Online Speech," *Harvard Law Review* 131, no. 6 (2018): 1599–670.

⁵⁹ Klonick, "The New Governors: The People, Rules, and Processes Governing Online Speech."

⁶⁰ Klonick, "The New Governors: The People, Rules, and Processes Governing Online Speech."

from posting it.⁶¹ In more serious cases, companies can also swiftly and comprehensively remove malicious content.⁶² Nevertheless, beyond reactionary measures, companies also implement preventive safeguards to reduce the spread of misinformation. Through partnerships with fact-checkers across languages, social media platforms ensure quick verification of content flagged as potentially misleading. Companies like TikTok have even launched media literacy campaigns across Europe to help users, especially younger audiences, critically assess content, identify misinformation, and access reliable sources.⁶³

Given the immense scale of misinformation on social media, platforms can't have human moderators screen content before it is published. Instead, most companies utilize automated moderation systems to identify misinformation *ex ante*, helping prevent its spread before it reaches the public.⁶⁴ However, automated moderation runs the risk of over-censorship, especially when a post blurs the line between what is acceptable and what is not. Only the worst forms of misinformation are immediately removed by the company, but the algorithm merely punishes the tolerable ones.

From these measures, this essay posits that there is an implicit hierarchy of misinformation in the eyes of social media companies. For outright harmful rhetoric likely to gain traction, social media companies may prioritise moderation through automated algorithms, ensuring that the post has zero reach. However, for issues that are more nuanced and contentious, the social media company may allow the content to be circulated, relying on human moderators to sift through it for misinformation.⁶⁵

Unfortunately, the current system of self-regulation still has significant limitations. Across the industry, platforms often interpret the hierarchy of misinformation differently, leading to varying degrees of content moderation. Some companies choose to err on the side of caution, removing most content that they deem as misleading or malicious. Other companies prefer to

⁶¹ Tarleton Gillespie, "Do Not Recommend? Reduction as a Form of Content Moderation," *Social Media + Society* 8, no. 3 (2022): 20563051221117552, <https://doi.org/10.1177/20563051221117552>.

⁶² Paul M. Barrett, "NYU Stern Center for Business & Human Rights: The Supreme Court Protects Social Media Platforms' Ability to Filter Out Harmful Content," accessed October 6, 2025, <https://bhr.stern.nyu.edu/quick-take/the-supreme-court-protects-social-media-platforms-ability-to-filter-out-harmful-content/>.

⁶³ TikTok, "Protecting the Integrity of TikTok during the Czech Parliamentary Elections," Newsroom | TikTok, accessed October 6, 2025, <https://newsroom.tiktok.com/protecting-the-integrity-of-tiktok-during-the-czech-parliamentary-elections?lang=en-150>.

⁶⁴ Alessia Zornetta, "Online misinformation: improving transparency in content moderation practices of social media companies," with Ignacio Cofone (Supervisor), McGill University, accessed October 21, 2025, <https://escholarship.mcgill.ca/concern/theses/rf55zd782>.

⁶⁵ Zornetta, "Online misinformation."

disregard controversial cases to prevent backlash from wrongly labelling content as misinformation.⁶⁶

Even within the platform, content moderation can be inconsistent. One striking example would be Twitter's treatment of US President Donald Trump's tweets. When President Trump made the controversial tweet, "when the looting starts, the shooting starts" in 2020, the company did not restrict it, and no further action was taken against his account. This matter has already raised questions about whether Trump should be prosecuted, given that public officials should be held to a higher standard. However, when another Twitter account copied the same message, the account was promptly suspended within three days.⁶⁷ This double standard reveals the lack of accountability in self-regulation: Social media companies often choose to regulate misinformation based on context, allowing it to continue spreading.

At the same time, all social media platforms are driven by profit. Social media companies earn revenue by maximising viewer engagement. Therefore, companies have an incentive to provide users with content they want to see, regardless of whether the information is true or false. As Harvard Law Professor Martha Minow stated, social media produces "echo chambers" with few opportunities to learn, understand, or believe what others are hearing as news. Social media algorithms are trained to feed us with information that reinforces what we already believe, thereby increasing our satisfaction by limiting our exposure to new, disruptive viewpoints.⁶⁸ Moreover, to maximise profit, these companies have an incentive to restrict content regulation. By adding more content to the platform, the company earns more from advertising revenue. Therefore, social media companies have a strong incentive to allow misinformation to spread, as content that gains traction generates advertising revenue.

Section 3.2: State-run regulation

State-run regulation, on the other hand, is vastly different. Instead of directly labelling information as misinformation, the government creates frameworks to regulate misinformation after social media companies have identified it. Unlike social media companies, governments do not have access to social media algorithms or back-end systems to remove misinformation from social media platforms. Thus, state-run regulation rarely addresses individual cases of misinformation, as it cannot directly control the

⁶⁶ Thomas James Vaughan Williams et al., "Policy vs Reality: Comparing the Policies of Social Media Sites and Users' Experiences, in the Context of Exposure to Extremist Content," *Behavioral Sciences of Terrorism and Political Aggression* 17, no. 1 (2025): 110–27, <https://doi.org/10.1080/19434472.2023.2195466>.

⁶⁷ Donie O'Sullivan, "Mark Zuckerberg Silent as Trump Uses Facebook and Instagram to Threaten 'Looting' Will Lead to 'Shooting' | CNN Business," CNN, May 29, 2020, <https://www.cnn.com/2020/05/29/tech/facebook-violence-trump>.

⁶⁸ Harvard Law School, "The Changing Ecosystem of News and Challenges for Freedom of the Press," *Harvard Law School*, n.d., accessed October 21, 2025, <https://hls.harvard.edu/bibliography/the-changing-ecosystem-of-news-and-challenges-for-freedom-of-the-press/>.

content available on its platforms.⁶⁹ Instead, most governments use a variety of methods to enforce regulations implemented by social media platforms.

Governments may enact legislation requiring social media companies to identify and remove misinformation promptly.⁷⁰ By holding social media companies liable for misinformation on their websites, state-run regulation compels them to implement procedural measures to address it. Alternatively, the government may issue advisories to social media companies to standardise best practices, thereby effectively limiting the spread of misinformation. For instance, the Ministry of Electronics and Information Technology (MeitY) in India issued an advisory in March 2024, stating that all social media platforms should inform users of the consequences of spreading misinformation. This includes removing content, suspending the user's right to post, or, in more serious cases, terminating the user's account.⁷¹

The government has an active incentive to implement state-run regulation for several reasons. From a principle perspective, the government has a duty to ensure democratic participation through all media, including cyberspace. Therefore, especially when misinformation directly targets other individuals and inhibits their ability to share their views freely online, the government has a moral responsibility to intervene.⁷² From a practical perspective, current attempts at self-regulation are insufficient. Even though each social media platform regulates misinformation independently, state-run regulation is still necessary to establish a common standard for platforms to work towards. This standardisation helps social media platforms to regulate misinformation within a set framework, rather than continuously treading the line between over-censorship and allowing misinformation to run rampant.⁷³ The government is also incentivised to regulate due to the unique context of the Russian-Ukrainian

⁶⁹ Ben Sperry, *Knowledge and Decisions in the Information Age: The Law & Economics of Regulating Misinformation on Social-Media Platforms* (International Center for Law & Economics, 2024), <https://laweconcenter.org/resources/knowledge-and-decisions-in-the-information-age-the-law-economics-of-regulating-misinformation-on-social-media-platforms/>.

⁷⁰ Howard Tumber and Silvio, "The Routledge Companion to Media Disinformation and Populism," Routledge & CRC Press, accessed October 21, 2025, <https://www.routledge.com/The-Routledge-Companion-to-Media-Disinformation-and-Populism/Tumber-Waisbord/p/book/9780367704919>.

⁷¹ Neeraj Soni, "Collaboration Between Government and Social Media Platforms to Counteract Online Misinformation," accessed October 21, 2025, <https://www.cyberpeace.org/resources/blogs/collaboration-between-government-and-social-media-platforms-to-counteract-online-misinformation>.

⁷² Sanders Robyn and Maciolek Ashleigh, "The Government Shouldn't Be Barred from Countering False Information | Brennan Center for Justice," September 12, 2025, <https://www.brennancenter.org/our-work/analysis-opinion/government-shouldnt-be-barred-countering-false-information>.

⁷³ Robyn and Ashleigh, "The Government Shouldn't Be Barred from Countering False Information | Brennan Center for Justice."

War. As the first “social media war” in history,⁷⁴ Both governments recognise the need to keep their citizens well-informed by combating the threat of misinformation. As much as the Russian-Ukrainian War is a physical battle of arms, it is also a battle of narratives.

State-run regulation is one of the most effective ways for both countries to tackle misinformation. This is because, unlike self-regulation, government policies are legally binding. For instance, when Russia first invaded Ukraine in 2022, the EU and the United Kingdom took drastic action by banning the Russian state broadcaster RT in their respective territories.⁷⁵ This limited European citizens’ exposure to Russian propaganda, thereby reducing the spread of pro-Russian misinformation. This demonstrates that the government has the legal mandate to regulate misinformation by any means necessary, even if that means complete censorship of information channels. When direct government intervention is not required, state-run regulation can also spur social media companies to act. Through advisories or direct legislation, state-run regulation acts as a strong signal of the government’s attitude towards this serious issue. This can push social media companies to review user content more regularly and thoroughly, in line with the guidelines and direction provided by the state-run regulation.⁷⁶

However, implementing broad-level state-run regulation alone is insufficient. By using frameworks to shape expectations, governments can only offer implementation suggestions.⁷⁷ This includes examples such as the European Commission’s voluntary Code of Practice on Disinformation, which comprises commitments from social media platforms to combat fake news.⁷⁸ While such frameworks are a step in the right direction, they do not include legally binding enforcement mechanisms. Thus, they are unable to hold social media companies to account effectively.

On the other hand, the government can wield its executive power and implement direct state-run regulation. This looks like the approach taken by Singapore through the 2019 Protection from Online Falsehoods and Manipulation Act. The Act allowed the government to determine what constitutes factual news and remove what it deemed “fake news.” Essentially, the Singapore government has been given full autonomy to control the flow of information

⁷⁴ Ciuriak, “Social Media Warfare Is Being Invented in Ukraine.”

⁷⁵ Jackson and Bateman, “Countering Disinformation Effectively: An Evidence-Based Policy Guide | Carnegie Endowment for International Peace.”

⁷⁶ Amélie Heldt, “Reading between the Lines and the Numbers: An Analysis of the First NetzDG Reports,” [info:eu-repo/semantics/article](https://info.eu-repo/semantics/article), Alexander von Humboldt Institute for Internet and Society gGmbH, June 12, 2019, <https://doi.org/10.14763/2019.2.1398>.

⁷⁷ Henman Paul, “The Big Question,” accessed October 21, 2025, <https://stories.uq.edu.au/contact-magazine/2019/big-question-social-media-regulation/>.

⁷⁸ European Commission, “The 2022 Code of Practice on Disinformation | Shaping Europe’s Digital Future,” accessed October 21, 2025, <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>.

through all public channels, thereby effectively regulating misinformation. Even within Singapore, this has raised concerns about whether the government is over-censoring free speech.

As noted by Chong Ja Ian, a political scientist working in Singapore, the law is “sweeping, broad, and contains vague wording.” This makes the Act a powerful tool for the Singapore government to control the dissemination of information, justifying the removal of content in the name of “fake news.”⁷⁹ In 2019, the Act was used to take down two Facebook posts by the Singapore Democratic Party (SDP) that were accused of spreading false information. The government presented a chart covering 2015-2018 to justify its decision. However, when the SDP protested that its statement was based on credible news articles, even taking the case to the Court of Appeal of Singapore, the Singapore government made no further clarification regarding its decision.⁸⁰

Notably, the SDP case also reveals an essential truth about the nature of direct state-run regulation. When the executive branch overextends its power, it is the role of judicial oversight to mediate that power. In this case, the SDP appealed against the executive decision, invoking the judiciary to assess the Minister’s interpretation of the content as false.⁸¹ This ensures that, even with direct state-run regulation, the government remains subject to legal review and accountability. While the SDP’s post was ultimately found to contain false statements of fact, the investigation nonetheless placed the executive branch’s decision under scrutiny.⁸² This judicial involvement helps to ensure state-run regulation remains accountable, preventing the government’s executive powers from going unchecked.

Most countries currently fall somewhere between the two opposing approaches: broad-level regulation and direct state-run regulation. Vietnam, for instance, has reached an agreement with Facebook to expedite content-removal requests from Vietnamese government ministries. Turkey has sought to compel social media platforms like Twitter and YouTube to

⁷⁹ Karishma Vaswani, *Concern over Singapore’s Anti-Fake News Law*, April 4, 2019, <https://www.bbc.com/news/business-47782470>.

⁸⁰ Robin Ramcharan, *Fake News and Elections in Southeast Asia: Impact on Democracy and Human Rights*, 1st ed. (Routledge, 2022), <https://www.routledge.com/Fake-News-and-Elections-in-Southeast-Asia-Impact-on-Democracy-and-Human-Rights/Gomez-Ramcharan/p/book/9781032379821>.

⁸¹ CNA, “Court of Appeal Partially Allows SDP’s POFMA Appeal in Landmark Judgment; Other Correction Directions Upheld,” CNA, accessed November 23, 2025, <https://www.channelnewsasia.com/singapore/pofma-sdp-appeal-partially-upheld-2230361>.

⁸² The Straits Times, “Court of Appeal Upholds Constitutionality of Pofma, Allows Part of SDP’s Appeal,” *The Straits Times* (Singapore), October 8, 2021, https://www.straitstimes.com/singapore/court-of-appeal-upholds-constitutionality-of-pofma-allows-part-of-sdps-appeal?utm_source=chatgpt.com.

remove defamatory content about Mustafa Kemal Atatürk, a renowned national figure often hailed as the founder of modern Turkey, but with little success.⁸³ Additionally, the United States is even proposing new legislation to crack down on social media bots and online political falsehoods.⁸⁴

In the case of the Russian-Ukrainian war, both broad-level and direct state-run regulation have been employed to combat misinformation. On a broader level, the EU's European External Action Service's East StratCom Task Force launched the EUvsDisinfo initiative, intending to issue guidelines and reports in response to the Russian government's disinformation campaign. This influences platform behaviour and public awareness, helping to reduce the spread of pro-Kremlin disinformation without legal compulsion.⁸⁵ At the same time, the EU has also taken concrete steps to remove pro-Kremlin misinformation sources directly. Following the invasion of Ukraine in 2022, the EU banned pro-Kremlin news outlets RT and Sputnik across all member states, regardless of content.⁸⁶

This further underscores concerns about state-run regulation. Beyond accounting for differing treatments towards the grey area of misinformation, we must also consider the possibility of a government outright censoring clearly valid information. Some sacrifices should be made when considering freedom of speech, but it must not become a weapon of oppression in the hands of the government. Going one step further, governments themselves can serve as sources of misinformation. According to Senator Rand Paul, the US government is the most significant source of misinformation during the COVID-19 pandemic.⁸⁷ In such cases, state-run regulation is ineffective, since the government can freely spread false narratives and conveniently decide what constitutes misinformation. State-run regulation provides a legitimate means for bad actors to censor information that contradicts their narrative, labeling it

⁸³ Daphne Keller, "Internet Platforms: Observations on Speech, Danger, and Money," Hoover Institution, accessed October 21, 2025, <https://www.hoover.org/research/internet-platforms-observations-speech-danger-and-money>.

⁸⁴ Ariel Shapiro, "Democratic Sen. Warner Has a New Policy Paper with Proposals to Regulate Big Tech Companies," CNBC, July 30, 2018, <https://www.cnbc.com/2018/07/30/sen-warner-proposes-20-ways-to-regulate-big-tech-and-radically-change.html>.

⁸⁵ EEAS, "EUvsDisinfo," accessed November 23, 2025, https://www.eeas.europa.eu/delegations/china/euvdisinfo_en?s=166.

⁸⁶ Gergely Ferenc Lendvai, "Media in War: An Overview of the European Restrictions on Russian Media," *European Papers - A Journal on Law and Integration* 2023 8, no. 3 (2024): 1235–45, <https://doi.org/10.15166/2499-8249/715>.

⁸⁷ Homeland Security, "Dr. Paul Stresses Importance of Investigating COVID-19 Origins to DHS Secretary," Committee on Homeland Security & Governmental Affairs, n.d., accessed December 11, 2025, <https://www.hsgac.senate.gov/media/rep/dr-paul-stresses-importance-of-investigating-covid-19-origins-to-dhs-secretary/>.

“misinformation.” This makes it difficult to hold governments accountable, especially when they have a vested interest in spreading misinformation.

SECTION 4: THE PATH FORWARD

With a better understanding of the misinformation crisis and the current global response, it is apparent that more must be done to prevent the spread of harmful misinformation. Although misinformation continues to plague the international community, current efforts to regulate misinformation has been disorganized and ineffective. Section 4 aims to outline a path forward and offer preliminary guidance on improving the current status quo.

Section 4.1: Lessons from the past

Based on Sections 1 to 3, it is apparent that the misinformation crisis is multifaceted. As evident in the Russian-Ukrainian war, misinformation has taken many forms, communication methods, and dissemination platforms. Ranging from doctored battlefield videos being uploaded onto Telegram to misleading political narratives spread through Reddit forums, there are endless ways for misinformation to manifest.⁸⁸ At the same time, misinformation spreads farther, deeper, and more widely than the truth. The culmination of these factors has created a systemic misinformation epidemic. Platforms have implemented algorithmic filters and fact-checkers, but they continue to lag behind the viral falsehoods that plague the internet.⁸⁹ State-run regulation is also being explored, but its heavy-handed nature makes it difficult for countries to implement effectively without being accused of abusing their executive power. To move forward, society cannot simply repeat ineffective policies. We must create a robust strategy that addresses the multifaceted nature of the issue.⁹⁰

There are two main ways to address the misinformation crisis related to the Russian-Ukrainian war. The first approach places the responsibility for misinformation squarely on platform users. This approach treats the spread of misinformation on social media as a reflection of society and, by extension, of human behaviour. During the Russian-Ukrainian war, some users shared misinformation to gain attention; others chose not to fact-check before sharing. Therefore, it does not matter if the system is changed, because people will continue to

⁸⁸ Clothilde Goujard, “YouTube’s Algorithm Pushes Hateful Content and Misinformation: Report,” *Politico*, July 7, 2021, <https://www.politico.eu/article/mozilla-firefox-report-youtube-algorithm-pushes-hateful-content-misinformation/>.

⁸⁹ Peter Dizikes, “Study: On Twitter, False News Travels Faster than True Stories,” *MIT News*, March 8, 2018, <https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308>.

⁹⁰ Robin Ramcharan, *Fake News and Elections in Southeast Asia: Impact on Democracy and Human Rights*, 1st ed. (Routledge, 2022), <https://www.routledge.com/Fake-News-and-Elections-in-Southeast-Asia-Impact-on-Democracy-and-Human-Rights/Gomez-Ramcharan/p/book/9781032379821>.

spread misinformation.⁹¹ The second approach involves treating social media as the source of the issue. Given that social media algorithms are built for sensationalist headlines and short-form content, it naturally favoured emotionally charged narratives about the death of innocent civilians or the pain of wounded soldiers, rather than logical reasoning and detailed explanations. As such, social media platforms are mainly responsible for creating a systemic framework that prioritises misinformation. Hence, without greater algorithmic transparency, efforts to change people's mindset will be ineffective in curbing the spread of misinformation.⁹²

However, neither approach alone tells the whole story. The former approach ignores algorithms that systematically recommend misinformation.⁹³ Without social media algorithms, malicious users would be able to amass only hundreds of views, thereby exponentially slowing the spread of misinformation. On the other hand, the latter approach fails to account for why individuals find misinformation memorable. Without users clicking on misinformation that pops up in their feeds, social media algorithms would also be unable to catalyze its spread.⁹⁴ As such, treating the misinformation crisis as purely a matter of human behaviour or algorithmic bias will lead to a myopic outcome. Instead, this essay purports that the proper solution could stem from a combination of both approaches. Even though some users have malicious intent, social media elevates these perspectives, pushing misinformation that is not representative of society at large.

As evident in Section 3, self-regulation and state-regulation are both limited in isolation but undeniably crucial for curbing the spread of misinformation. To address this issue, a hybrid approach that accounts for the unique strengths and limitations of both actors is necessary. Self-regulation can be used as the first line of defence against misinformation. This is because social media companies have the exclusive ability to implement technical interventions, such as deprioritising misleading posts and flagging suspicious content.⁹⁵ With complete access to the algorithm, social media companies will be able to adapt misinformation regulations to overcome platform-specific challenges swiftly. This can be complemented using state-run regulation in more critical situations, where the potential for public harm is immediate and severe. In these cases, the government must rely on its executive powers to regulate misinformation related to high-stakes issues, such as false claims regarding public health data or emergency response news. Nevertheless, the government should also exercise restraint in cases involving national sensitivities. As evidenced through the case study of Mustafa Kemal Atatürk in Turkey, governments must not compel social media companies to remove uncomfortable content from

⁹¹ Joshua A. Tucker et al., "From Liberation to Turmoil: Social Media And Democracy," *Journal of Democracy* 28, no. 4 (2017): 46–59, <https://doi.org/10.1353/jod.2017.0064>.

⁹² Sierra L. Lawson and Steven W. Ramey, "Sourcing Stereotypes: Constructing and Challenging Simplified Knowledge," *Culture and Religion* 19, no. 4 (2018): 416–34, <https://doi.org/10.1080/14755610.2018.1505759>.

⁹³ Goujard, "YouTube's Algorithm Pushes Hateful Content and Misinformation: Report."

⁹⁴ Franziska Zimmer et al., *Fake News in Social Media: Bad Algorithms or Biased Users?*, Korea Institute of Science and Technology Information, June 30, 2019, <https://doi.org/10.1633/JISTaP.2019.7.2.4>.

⁹⁵ Corinne Tan, "Regulating Disinformation on Twitter and Facebook," *Griffith Law Review* 31, no. 4 (2022): 513–36, <https://doi.org/10.1080/10383441.2022.2138140>.

their platforms. Instead, it must be keenly aware of the risk that state intervention crosses the line into authoritarian overreach.⁹⁶

However, the isolated efforts of individual countries may be insufficient to solve the misinformation crisis. With its international reach and extensive viewership, social media platforms have effectively become superstate actors who transcend geographical boundaries.⁹⁷ As such, no single nation can exercise complete control over content posted on social media platforms, nor can it unilaterally enforce accountability on platforms beyond its jurisdiction. To address this power imbalance, the next logical step is for the international community to move towards multilateral cooperation and international treaties. By establishing global standards, countries can create a standardised regulatory framework that all social media companies must adhere to. This allows for more straightforward guidelines on the scope of regulation and on how to hold social media companies accountable through self-regulation agreements and state-regulated directives. Even if multilateral forums seem unproductive, they still offer an essential avenue for discourse.⁹⁸ This allows countries to voice regulation concerns in a cordial setting, preventing disagreements from escalating into global tensions or censorship wars.

Section 4.2: Envisaging the future

Going one step further, this paper argues that a long-term solution requires not just an effective response to misinformation after it has spread, but also an emphasis on preventing its spread on social media platforms in the first place. This is because it is exponentially harder to reverse the effects of misinformation after it has gone viral. Therefore, fact-checking and takedown efforts, while necessary, are too late to fully neutralise the harms caused by misinformation. Incorporating a preventive framework in the future can reduce the burden on reactive regulation systems.⁹⁹ There are a few key characteristics of this proposed long-term solution.

Firstly, social media companies must prioritise accuracy over sensationalism. While it is unavoidable for social media algorithms to promote content based on popularity, companies must also take steps to verify content accuracy. For instance, social media companies could implement a tiered verification system that involves expedited fact-checking for viral posts. This means that, even as social media algorithms prioritise engaging content, there will be a parallel mechanism

⁹⁶ Tech Against Terrorism, “The Online Regulation Series,” accessed November 7, 2025, <https://techagainstterrorism.org/news/2020/10/23/the-online-regulation-series-turkey>.

⁹⁷ Bandopadhyaya Suvojit, “Transcendence through Social Media,” *Journal of Media and Communication Studies* 8, no. 3 (2016): 25–30, <https://doi.org/10.5897/JMCS2015.0469>.

⁹⁸ UNESCO, “Online Disinformation : UNESCO Unveils Action Plan to Regulate Social,” accessed November 7, 2025, <https://www.unesco.org/en/articles/online-disinformation-unesco-unveils-action-plan-regulate-social-media-platforms>.

⁹⁹ Kate Klonick, “The New Governors: The People, Rules, and Processes Governing Online Speech,” *Harvard Law Review* 131, no. 6 (2018): 1599–670.

to ensure content accuracy. This prevents misinformation from spreading by ensuring that a heightened sense of responsibility over content accuracy accompanies the extensive reach of social media.¹⁰⁰

Secondly, users must evolve from passive consumers to active stakeholders. While the government and social media companies may enact joint regulation through state-level and international-level treaties, users must also collaborate with both stakeholders. On the one hand, social media users must work with companies to effectively moderate social media platforms. One good example of this is TikTok's efforts to collaborate with researchers and fact-checkers. By partnering with users across all official EU languages, the company ensures a swift response to international misinformation reports.¹⁰¹ On the other hand, social media users must also be receptive towards the government's misinformation education efforts. Through which, users will be incentivised to recognise and call out potential instances of misinformation.

Lastly, this paper also calls for greater transparency and accountability across all stakeholders. At its core, the misinformation crisis cannot be solved without genuine collaboration between stakeholders. Social media companies must strive for complete transparency about their algorithms. Through regular accountability reports, third parties can hold social media companies to account, ensuring that they take adequate measures to combat harmful content on their platforms. Additionally, the need for governmental transparency is two-fold. While the government should be transparent about how it chooses to regulate and enforce platform obligations, it must also be transparent about its decision-making process regarding government censorship.¹⁰²

In conclusion, the long-term solution described in this section is merely a guide to the path forward. It serves as a direction for society to take, given the current landscape of the misinformation crisis. More importantly, however, society must remain open to new developments regarding the misinformation crisis while remaining unwavering in the need for regulation. Ultimately, the fight against misinformation cannot be merely a series of reactive measures against falsehoods. Instead, the collective efforts of all stakeholders must be geared towards cultivating a proactive digital environment that calls out misinformation and embraces the truth.

¹⁰⁰ Otávio Vinhas and Marco Bastos, "Fact-Checking Misinformation: Eight Notes on Consensus Reality," *Journalism Studies* 23, no. 4 (2022): 448–68, <https://doi.org/10.1080/1461670X.2022.2031259>.

¹⁰¹ Caroline Greer, "TikTok's Continued Commitment to Combating Disinformation Under the EU Code of Practice," September 24, 2024, <https://newsroom.tiktok.com/tiktoks-continued-commitment-to-combating-disinformation-under-the-eu-code-of-practice?lang=en-150>.

¹⁰² Mark MacCarthy, "Transparency Is Essential for Effective Social Media Regulation," *Brookings*, January 11, 2022, <https://www.brookings.edu/articles/transparency-is-essential-for-effective-social-media-regulation/>.

The Link Between Solar UV Radiation and the Development of Skin Cancer Through DNA Damage By Emi Kariqi

Abstract

Solar UV radiation is a major health risk to humans, as it can lead to the development of skin cancer. This occurs mainly through interactions with DNA, but also involves changes at the immunological and clinical level. Upon UV exposure (UV-A and UV-B), DNA is damaged by the formation of oxidative lesions and photoproducts. This damage can be resolved through different repair pathways and cellular responses, but in the cases that it is not, it can result in mutations. Accumulation of these mutations in tumor suppressor or growth-stimulatory genes then leads to skin cancer. At the immunological level, UV radiation depletes tumor-detecting cells and increases production of factors that inhibit the immune system. Meanwhile at the clinical level, sunburns develop, increasing oxidative DNA damage, while tanning is activated as a defense mechanism against UV. This paper will act as a cohesive review on the relationship between UV exposure and skin cancer, highlighting all core processes while addressing overlooked aspects of the topic such as the role of UV-A radiation and the overlap between UV-A and UV-B mechanisms. The overarching aim is to underline the importance of sun protection and avoiding practices such as excessive tanning.

Introduction

Skin cancer is one of the most common cancers in the world, and its incidence is continuously increasing, reaching 1.5 million cases worldwide in 2022 (WHO). It is categorized into melanoma, and non-melanoma skin cancer (NMSC) (Roky et al., 2024). Melanoma is the most aggressive type, as it tends to spread to other areas of the body. Melanoma develops in melanocytes, the cells that produce the pigment melanin, which gives skin its color and protects it from solar radiation. NMSC, on the other hand, develops in keratinocytes, the principal cell type in the epidermis. NMSC is further divided into basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). For all three main types of skin cancer, exposure to solar ultraviolet radiation is a major risk factor (Roky et al., 2024; Cleveland Clinic, 2021). In 2022, it was reported that 83% of melanoma cases worldwide were directly linked to ultraviolet exposure, demonstrating the strong connection between sunlight and skin cancer (Langselius et al., 2025).

Sunlight comprises three types of electromagnetic radiation: ultraviolet (UV), visible, and infrared. UV radiation makes up around 40% of sunlight and contains the highest energies of light among the three regions. This means that it poses the greatest risk to human health (Baron and Suggs, 2014). UV spans the wavelengths 200 to 400 nm, and is divided into three categories: UV-A (320-400 nm), UV-B (280-320 nm), and UV-C (200-280 nm). UV-C is blocked by the ozone layer from penetrating the earth's surface (**Fig. 1**) (Matsumura and Ananthaswamy, 2019). In cases of exposure from man-made sources, it minimally penetrates the skin due to its shorter wavelengths, which can cause acute skin burns, but is not associated with long term damage or skin cancer risk (Health Center for Devices and Radiological, 2020). UV-A makes up the

majority of solar UV radiation (95%) (Baron and Suggs, 2014). It is considered weakly carcinogenic, as its effect on DNA is indirect, yet plays a major role in the development of melanoma (Ridley et al., 2009; R  nger, 1999). Due to its longer wavelengths, it penetrates deep into the dermis (**Fig. 1**), where it is absorbed by molecular chromophores. This induces oxidation reactions that generate reactive oxygen species (ROS), which in turn damage cellular components, including DNA (Baron and Suggs, 2014). UV-B, which makes up around 5% of solar UV radiation, is the most carcinogenic. It penetrates the epidermis (**Fig. 1**), where it is absorbed by DNA, causing photoproducts to form between adjacent nucleotide bases (Baron and Suggs, 2014; Ravanat et al., 2001). It is important to note that even though distinctions are made between the types of damage induced by UV-A and UV-B, these differences are not absolute. Oxidative damage can occur in the UV-B range and photoproducts can form in the UV-A range (R  nger, 1999).

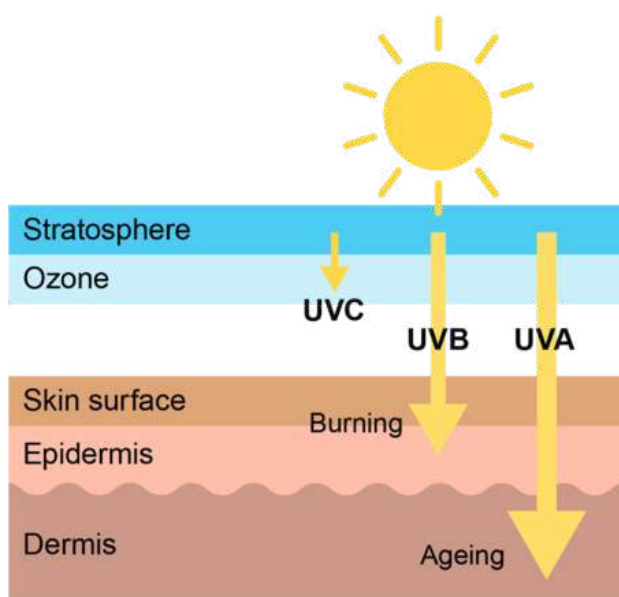


Figure 1. The wavelengths of UV radiation and their penetration depth (Cancer Research UK)

In both UV-A and UV-B exposure, once photons are absorbed, a photochemical reaction occurs causing the above-detailed cellular changes (Baron and Suggs, 2014). DNA repair pathways and cellular stress responses are then activated to resolve these changes, however mutations can still occur in certain genes, driving cancer. Additionally, UV exposure can independently suppress the immune system, making it easier for cancer cells to evade detection (Matsumura and Ananthaswamy, 2004).

As these changes occur at the molecular level, on the clinical level, the skin responds to solar radiation by burning (erythema) or tanning. Erythema is an inflammatory reaction that involves the vasodilation of blood vessels, and it has been linked to increased risk of skin cancer (Baron and Suggs, 2014; Wu et al., 2016). The primary cause is direct damage to DNA by UV-B,

and is more common in fair-skinned individuals due to the lack of melanin to protect the skin from this damage. On the other hand, tanning, which is the pigmentation of skin after UV exposure, protects from further UV damage. These protective properties arise from the abundance of melanin, which absorbs UV radiation and prevents it from penetrating the skin (Baron and Suggs, 2014).

The purpose of this article is to comprehensively outline the chain of events from initial UV exposure to the development of skin cancer, across the molecular, immunological, and clinical level. By doing so, it aims to inform readers about the risks of UV-induced skin cancer. Furthermore, because the role of UV-A in skin cancer development has often been overlooked by past studies and sunscreen manufacturers, this article aims to address it in full and highlight its overlap with UV-B radiation (Rünger, 1999).

1. Effect of UV radiation on the molecular level

1.1. DNA damage

When skin is exposed to UV-A and UV-B radiation, epidermal DNA is damaged through different pathways. Each process begins with photons being absorbed by certain molecular components in the skin called chromophores. This induces a photochemical reaction during which chromophores move to an excited state, causing them to undergo chemical changes, transfer energy to other molecules, or release extra energy as heat or light (Baron and Suggs, 2014; Rabe et al., 2006). The result of this is DNA lesions, (alterations to DNA's bases or sugar-phosphate backbone), which if not properly repaired, can give rise to harmful mutations.

1.1a. UV-A-induced damage

In the case of UV-A radiation, its chromophores have not yet been determined, but they are thought to be lipids, proteins, and a few other molecules. Altogether, they are classified as endogenous photosensitizers. DNA is not considered a chromophore for UV-A because it minimally absorbs radiation in this range of wavelengths (Baron and Suggs, 2014). When the above-mentioned photosensitizers absorb UV-A, they cause indirect damage to DNA through photooxidation reactions. The main target of these reactions is guanine, as it has the lowest ionization energy out of the four DNA bases. In type I photooxidation reactions, a guanine radical cation is formed through either electron abstraction or hole transfer (Ravanat et al., 2001). Electron abstraction occurs when the excited chromophore takes an electron from guanine. On the other hand, hole transfer happens when a distant DNA base that is not guanine gets oxidized by absorbing UV radiation, and its charge (or hole as it is referred to) migrates to a guanine base (Hall et al., 1996). The newly formed guanine radical cation then undergoes one of the following processes: hydration followed by oxidation, which gives 8-oxo-7,8-dihydroguanine (8-oxoGua); or hydration followed by reduction, which gives 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyGua) (**Fig. 2**). As for type II photooxidation, the mechanism involves the excited

chromophore transferring energy to molecular oxygen to produce singlet molecular oxygen ($^1\text{O}_2$). This product then oxidizes electron-rich molecules such as guanine to form 8-oxo-7,8-dihydroguanine (8-oxoGua) (Ravanat et al., 2001). Pouget and Douki (2000) demonstrated, using HPLC-EC and HPLC/GC-MS systems, that 8-oxoGua is the major product of UV-A exposure, a finding of special importance given its well-established mutagenic potential (Baron and Suggs, 2014). In the event these lesions fail to be repaired by cellular mechanisms, cancer-causing mutations may occur.

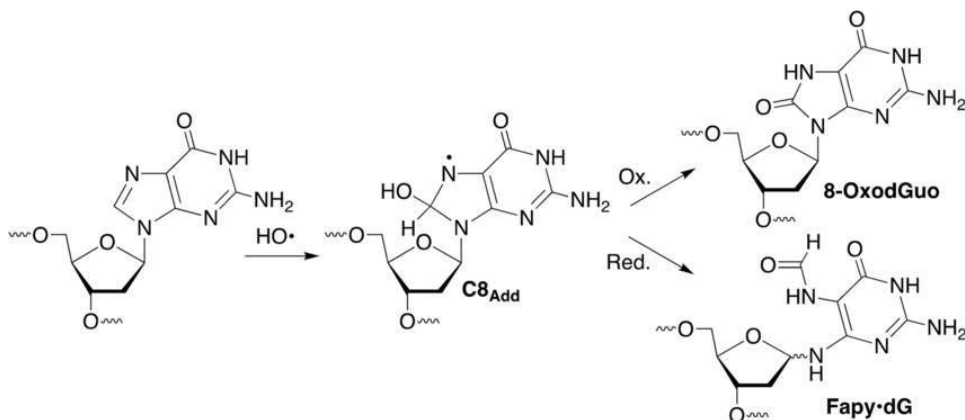


Figure 2. The production of oxidative lesions via type I photooxidation reactions (Yang et al., p. 5441)

Besides these lesions, UV-A can induce DNA damage through the production of hydroxyl radicals. In type I photooxidation reactions, once the guanine radical cation is formed, the reduced chromophore (which has gained the electron lost from guanine) reacts with oxygen to give a superoxide radical anion ($\text{O}_2^{\cdot-}$). Although this product is mostly unreactive, it can be converted to the highly reactive hydroxyl radical ($\cdot\text{OH}$), via the decomposition of hydrogen peroxide by ferrous ions (the Fenton reaction) (Pouget et al., 2000; John, 2022). $\cdot\text{OH}$ further reacts with DNA, causing DNA strand breaks, modified bases, abasic sites, sugar alterations, and DNA-protein cross-links (Pouget et al., 2000). Additionally, UV-A is also capable of producing cyclobutane pyrimidine dimers (CPDs), which are mainly a product of UV-B exposure, as explained below (Baron and Suggs, 2014).

1.1b. UV-B-induced damage

UV-B radiation has DNA as its chromophore because it strongly absorbs in this range of wavelengths. At the moment of absorption, bulky photoproducts form between two adjacent nucleotide bases, mainly cyclobutane pyrimidine dimers, and pyrimidine (6-4) photoproducts (Mullenders, 2018). These lesions occur mostly in pyrimidines (cytosine and thymine) and to a lesser extent in purines (adenine and guanine) (Ravanat et al., 2001). They both involve the C5-C6 double bonds of DNA bases, which absorb UV radiation most efficiently (Baron and

Suggs, 2014). CPDs are formed through the cycloaddition of C5-C6 double bonds of two adjacent pyrimidine bases (**Fig. 3**). This lesion has six possible stereoisomers, arising from the cis/trans configurations of the pyrimidines relative to the cyclobutane ring, as well as the syn/anti configurations relative to the position of the C5-C6 bond (Cadet et al., 1985). These isomers may differ in how strongly they damage DNA, how efficiently they are repaired, or what mutations they induce; however there is insufficient research to confirm or characterize these differences. In DNA and oligonucleotides only syn isomers exist due to steric constraints, with the cis-syn form being produced predominantly over the trans-syn form.

The second most common UV-B-induced lesions, pyrimidine (6-4) photoproducts, form through a two step process. First, the C5-C6 double bond of the 5'-end pyrimidine undergoes cycloaddition with the C4 carbonyl group of the 3'-end pyrimidine, forming an unstable oxetane intermediate. An azetidine intermediate may be formed if the 3'-end base is a cytosine and its 4-imino function bonds with the 5'-end pyrimidine through cycloaddition. Both the oxetane and the azetidine can rearrange spontaneously to give pyrimidine (6-4) photoproducts (**Fig. 3**). Additionally, UV-B irradiated DNA forms secondary photoproducts called related Dewar valence isomers, which rise from (6-4) photoproducts through further sunlight exposure (**Fig. 3**). In total, UV-B induces twelve types of photoproducts in DNA, depending on the lesion type (CPDs, pyrimidine (6-4) photoproducts, related Dewar valence isomers) and the combination of pyrimidines (TT, CC, TC, CT) (Ravanat et al., 2001).

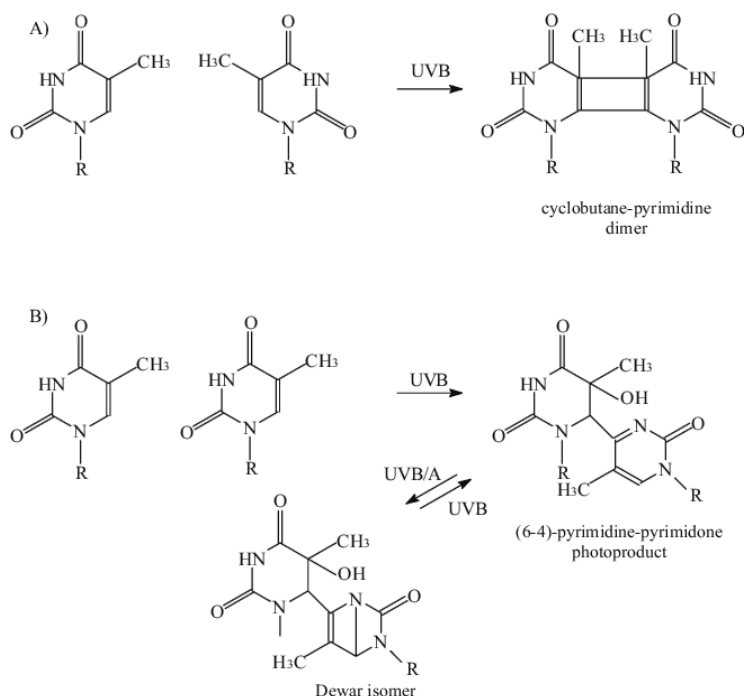


Figure 3. The production of cyclobutane pyrimidine dimers, pyrimidine (6-4) photoproducts, and Dewar valence isomers (Svobodova et al., p. 28)

These lesions can lead to the development of cancerous mutations, as a result of errors in DNA repair. Out of all the lesions, CPDs have the highest abundance, while Dewar valence isomers are produced at much lower amounts, as demonstrated by Douki and Cadet (2001) through HPLC-MS/MS analyses. This is particularly important as CPDs are the main contributors to mutations in mammals (Tornaletti and Pfeifer, 1996). Additionally the study shows that TT and TC generate the highest amount of lesions, while CT and TC generate significantly less (Douki and Cadet, 2001).

Both UV-A and UV-B exposure damage DNA, indirectly and directly, by producing mutagenic lesions that can affect core processes such as DNA repair and replication (Iordanou and Berneburg, 2010). To protect DNA and prevent mutations, all cells are equipped with repair pathways, which work to remove the harmful lesions (Matsumura and Ananthaswamy, 2004).

1.2. DNA repair pathways

Depending on the type of lesion, one or more DNA repair mechanisms are activated following DNA damage (Matsumura and Ananthaswamy, 2004). For oxidative damage, base excision repair (BER) and single strand break repair (SSBR) are used. For bulky photoproducts, cells utilize a comparatively more complex pathway called nucleotide excision repair (NER) (Ridley et al., 2009).

1.2a. Base excision repair and single strand break repair

Base excision repair (BER) is initiated by a glycosylase which recognizes small base lesions that do not significantly disrupt the DNA helix. There are 11 identified mammalian glycosylases and each is responsible for recognizing specific related lesions (Krokan and Bjoras, 2013). After identification of the lesion, glycosylases break the glycosidic bond between the damaged DNA base and the sugar-phosphate backbone, creating what is known as an abasic (AP) site. AP endonuclease 1 (APE1) enzymes or bifunctional glycosylases then cleave the DNA backbone at the AP site, creating a one nucleotide gap bordered by 3'-hydroxyl and 5'-deoxyribosephosphate (5'-dRP). DNA polymerase β removes 5'-dRP residues and places the correct nucleotide in the gap. Finally, DNA ligase III α together with X-ray cross-complementing protein 1 (XRCC1), binds the new nucleotide to the sugar-phosphate backbone. This is the predominant pathway of BER, which is known as short patch BER (**Fig. 4**). In cases where 5'-dRP residues are modified and cannot be removed by DNA polymerase β , long patch BER is used. In this pathway, DNA polymerase δ or ϵ acts by placing multiple nucleotides in the single nucleotide gap, creating a DNA flap structure. This structure is cut and removed by flap endonuclease-1 (FEN-1), then DNA Ligase I acts to seal the sugar-phosphate backbone (**Fig. 4**) (Carter and Parsons, 2016).

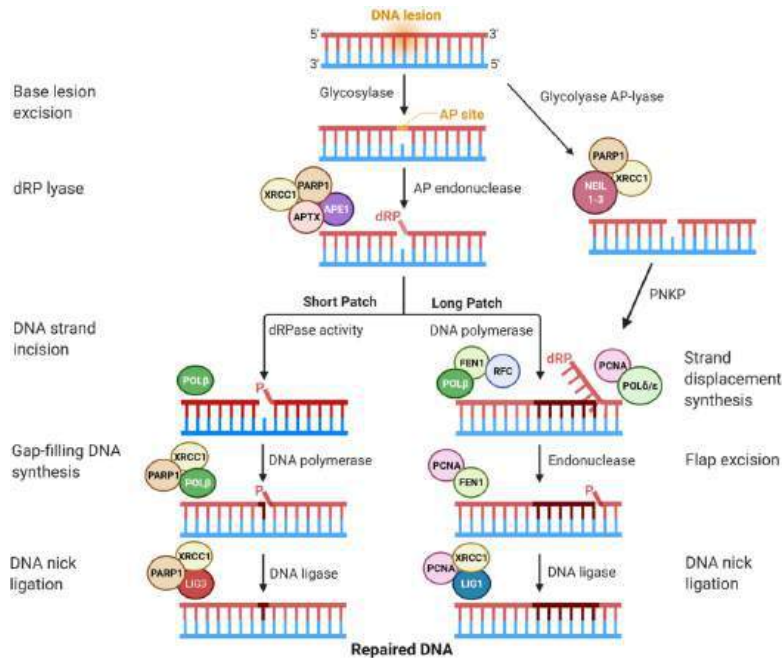


Figure 4. The base excision repair (BER) pathway (Hindi et al.)

Studies by Jiang et al. (2009) and Starcevic (2017) have shown that alterations in BER proteins, such as single nucleotide polymorphisms, are linked to increased risk of cancer. This demonstrates how crucial BER is in the repair of oxidative lesions and the prevention of further damage.

Another pathway, single strand break repair (SSBR), works to repair single strand breaks (SSBs) caused by hydroxyl radicals. A group of enzymes called poly (ADP-ribose) polymerases (PARPs), recognize SSBs and bind to them, subsequently becoming catalytically activated. Upon activation, they transfer ADP-ribose to themselves and nearby proteins. This process unravels chromatin, allowing repair proteins to access the damaged DNA site. Additionally, the attachment of ADP-ribose enables PARP enzymes to recruit XRCC1 protein complexes, which contain most of the enzymes necessary for SSBR. Polynucleotide kinase phosphatase (PNKP) and aprataxin (APTX) convert the ends of SSBs to 3'-hydroxyl and 5'-deoxyribosephosphate (5'-dRP), which are required for DNA synthesis and ligation. Then, the process continues through either short patch or long patch BER, with DNA polymerases placing the correct nucleotides, FEN-1 cutting flap structures if present, and DNA ligases sealing the sugar-phosphate backbone. Unlike in other DNA repair pathways, mutations in SSBR enzymes and proteins have been linked to neurodegenerative diseases rather than cancer (Caldecott, 2022). This could suggest that SSBR is not a major pathway for repairing UV-A-induced damage, or that UV-A induces relatively few SSBs.

1.2b. Nucleotide excision repair

Nucleotide excision repair proceeds through two subpathways: transcription-coupled repair (TCR) and global genome repair (GGR). TCR is the faster process and occurs in actively transcribing genes that are damaged by UV radiation. It uses RNA polymerase II to identify the damaged site, with cockayne syndrome group A and B proteins (CSA and CSB) aiding the

process. In contrast, GGR is slower and acts on the remaining UV-damaged genes that are not repaired by TCR. It uses the XPC/HHR23B protein complex to identify the damaged site. After the initial damage recognition, both mechanisms proceed in the same way. Two helicase subunits, xeroderma pigmentosum type B and D (XPB and XPD), hydrolyze ATP and use the energy to unwind the DNA double helix around the lesion, creating a “repair bubble” that is around 20 nucleotides long. Other repair proteins then cut the DNA strands, making the 3’ incision first, then the 5’ incision. The excised DNA part dissociates, while DNA synthesis proteins and DNA polymerase δ or ϵ are recruited to fill in the gap with the correct nucleotides. Finally, DNA Ligase I connects the new nucleotides to the sugar-phosphate backbone (**Fig. 5**). While NER is best known for the repair of photoproducts, it also acts as a backup mechanism for the repair of oxidative lesions (Reardon and Sancar, 2005).

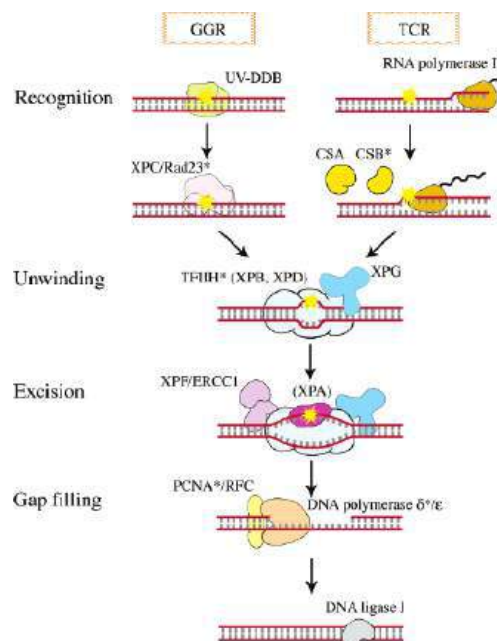


Figure 5. The nucleotide excision repair (NER) pathway (Kimura and Sakaguchi, p. 758)

Many studies have linked inherited defects in NER to diseases such as xeroderma pigmentosum (XP), Cockayne syndrome (CS), UV sensitive syndrome (UVsS) and the photosensitive form of trichothiodystrophy (TTD). All of these diseases make the carriers more UV sensitive, but XP especially makes patients 1000 times more susceptible to UV-induced skin cancer (Mullenders, 2018). The fact that the dysfunction of NER has such a strong correlation with skin cancer risk emphasizes how crucial this pathway is in preventing cancerous mutations.

Overall, BER, and NER are key players in preventing skin cancer following UV-A and UV-B-induced damage. However, damage can persist even after these pathways, in which case the cell activates additional defense mechanisms (Matsumura and Ananthaswamy, 2004).

1.3. Cell stress responses

The cell responds to unrepaired UV-induced damage through cell cycle arrest, apoptosis, or antioxidant activity. In the first two mechanisms, the p53 tumor suppressor protein plays a major role. Upon UV exposure, the cell increases the expression of TP53, the gene that encodes for p53 (Matsumura and Ananthaswamy, 2004; Hernández Borrero and El-Deiry, 1876). Meanwhile the amino acids Serine 15 and Serine 20 are phosphorylated to activate p53, which in turn induces the expression of the p21 protein. p21 then competitively inhibits cyclin-dependent kinases (CDKs), leading to cell cycle arrest at the G1 phase. Stopping the cell cycle at this checkpoint allows time for DNA damage to be repaired before replication, therefore reducing the risk of propagating mutations. However, the UV-induced damage can be too severe to repair, and in such cases, the cell undergoes apoptosis (cell death) (Matsumura and Ananthaswamy, 2004).

Cell apoptosis is a tightly regulated process that occurs in apoptotic keratinocytes, also called “sunburn cells”. Keratinocytes are affected because they proliferate at higher rates than melanocytes, meaning they are more likely to be targeted for elimination during DNA replication (Baron and Suggs, 2014). Apoptosis is induced by the p53 protein, which increases the expression of apoptosis-promoting genes (Bax, Fas/Apo1), and decreases the expression of apoptosis-suppressing genes (Bcl-2) (Matsumura and Ananthaswamy, 2004). The Fas gene is especially important as studies using UV irradiated mice have shown that interactions between the gene and its ligand (Fas-L) are crucial in forming apoptotic cells (Hill et al., 1999). Upon apoptosis, the cell undergoes morphological changes, including shrinkage, membrane blebbing, DNA fragmentation, and the formation of a pyknotic nucleus. The remaining apoptotic body (the sunburn cell) is consumed by macrophages, allowing for it to be removed without harming the remaining tissue (Baron and Suggs, 2014).

Besides cell cycle arrest and apoptosis, the cell responds to UV exposure using antioxidants, which prevent ROS molecules from inducing oxidative DNA lesions. Numerous antioxidants exist, with each having different roles in protecting the cell. For instance, superoxide dismutase (SOD) catalyzes the conversion of superoxide into hydrogen peroxide, which is a crucial step in the removal of ROS, while glutathione tripeptide (GSH) detoxifies ROS due to its thiol groups (Ridley et al., 2009). Other antioxidants, such as ascorbic acid, have been shown to reduce mutation frequency after UVA irradiation (Besaratnia et al., 2007). However, a major limitation is that antioxidant activity is suppressed by UV radiation, thus weakening the cell's defenses against ROS and increasing the prevalence of oxidative lesions that can transform into mutations (Ridley et al., 2009).

1.4. Mutations

If UV-induced damage remains even after DNA repair and cell stress responses, it can give rise to permanent mutations, known as “UV signature mutations”. CPDs and pyrimidine (6-4) photoproducts (UV-B-induced lesions) frequently lead to C→T or CC→TT transitions. This trend is explained by the “A” rule, which states that when DNA polymerase encounters a non-instructional site, meaning a site that it cannot interpret, it inserts adenine bases by default. For example, in CC CPDs, DNA polymerase places two adenines opposite to the lesion instead of the correct guanines. The adenines are then paired with two thymines during the next round of

DNA replication, leading to CC→TT transitions. Similarly, in TC pyrimidine (6-4) photoproducts, the cytosine is read as a non-instructional base and it is paired with adenine, leading to C→T transitions (Matsumura and Ananthaswamy, 2004). In contrast, 8-oxoGua (UV-A-induced lesion) is commonly associated with G→T transitions. The principle is the same, but since oxidative lesions mainly occur in guanines, it is the guanine base that is misread as a non-instructional site and paired with adenine instead of cytosine (Rünger, 1999). Oxidative lesions can also lead to C→T transitions, however this has only been reported in bacteria and G→T remains as their main transition type (Reid and Loeb, 1993).

These mutations play a key role in driving skin cancer development, as they can cause the inactivation of tumor suppressor genes, or the expression of growth-stimulatory genes, called proto-oncogenes. Depending on where they occur in the genome, they give rise to different forms of skin cancer (Matsumura and Ananthaswamy, 2004).

1.4a. Mutations that drive NMSC

Non-melanoma skin cancers develop in keratinocytes as a result of cumulative UV exposure (Baron and Suggs, 2014). They are strongly linked to mutations in the p53 gene, which occur in 50-90% of NMSCs, and 50% of all human cancers. Normally, p53 is responsible for protecting the cell from extracellular stress (such as UV radiation) through cell cycle arrest and apoptotic pathways. When mutated, this gene loses its protective function, allowing damaged cells to survive and accumulate further mutations. Most p53 mutations are CC→TT transitions, indicating that UV-B likely plays a major role in NMSC development (Matsumura and Ananthaswamy, 2004). This is further supported by mouse models, which show increasingly higher rates of SCC (a form of NMSC), at shorter UV wavelengths (Rünger, 1999).

Other genes related to NMSCs include CDKN2A, Patched and Ras. CDKN2A is a tumor suppressor that encodes two proteins: p16INK4a, which inhibits cyclin-dependent kinases, stopping the cell cycle at the G1 phase; and p19ARF which inhibits p53 degradation (Matsumura and Ananthaswamy, 2004). CDKN2A mutations have been linked to SCC development, with a study by Soufir (1999) showing that in 20 selected cases of SCC, all of them involved mutations leading to p16INK4a and p19ARF inactivation. Additionally, 75% of the SCC cases exhibited UV signature mutations, confirming their UV origin.

Patched is another tumor suppressor gene, which is strongly associated with BCCs (a form of NMSC). Unlike p53 and CDKN2A, only a portion of patched mutations display the UV signature, suggesting additional mutagenic pathways alongside UV exposure. Finally, Ras is a proto-oncogene that, when activated, induces benign epidermal tumors or malignant conversions (Matsumura and Ananthaswamy, 2004). Studies using mouse models have detected Ras mutations in 20-40% of UV-induced NMSCs, however this percentage is significantly lower in humans (Nishigori et al., 1994). Additionally, Ras mutations derive from oxidative DNA damage such as 8-oxoGua, instead of photoproducts (Ito et al., 2025). This suggests that UV-A also plays a role in NMSC development.

1.4b. Mutations that drive melanoma

Melanoma skin cancer develops in melanocytes as a result of both genetic predispositions and environmental risk factors. Intense, intermittent UV exposure plays a particularly major role. Familial melanoma is strongly associated with mutations in the CDKN2A (p16) gene, which is referred to as a melanoma susceptibility gene. In contrast, sporadic melanomas show a low frequency of CDKN2A mutations (0-26%). A tumor suppressor gene in chromosome 1p36 is also involved in the development of melanoma and other skin cancers, although it has not been identified yet. Regarding the origin of mutations that drive melanoma, only around half of them exhibit the UV signature, which suggests additional mutagenic pathways (Matsumura and Ananthaswamy, 2004).

Not much is known about which range of UV induces melanoma; however a study using the hybrid fish *Xiphophorus* demonstrated that UV-A was largely more effective than UV-B in inducing melanoma (Setlow et al., 1993). In addition, users of tanning salons, which emit UV-A radiation, have shown up to an 8-fold increase in melanoma risk (Rünger, 1999). This demonstrates that UV-A likely plays a significant role in melanoma development.

2. Effect of UV radiation on the immunological level

2.1. Immunosuppression

The skin contains a group of specialized immune components that are collectively called skin-associated lymphoid tissues (SALT). These include Langerhans cells (LC), keratinocytes, and regional draining lymph nodes. LCs are antigen presenting cells, meaning they capture foreign antigens and show them to T lymphocytes. Keratinocytes produce cytokines, which contribute to immune responses, while lymph nodes filter lymph fluid from germs, damaged cells, and cancer cells (Matsumura and Ananthaswamy, 2004; Cleveland Clinic, 2022). Altogether, these components play a critical role in tumor surveillance.

However, the SALT system is disrupted by UV radiation, which induces local and systemic immunosuppression (Rabe et al., 2006). UV suppresses contact hypersensitivity (CHS) and delayed-type hypersensitivity (DHS), two crucial immunologic responses mediated by T cells (Matsumura and Ananthaswamy, 2004; L.A. Burns-Naas, 2014). Upon UV exposure, LCs undergo functional and structural changes that lead to their depletion from the epidermis, while keratinocytes produce immunosuppressive cytokines such as interleukin-10 (IL-10) (Rabe et al., 2006; Matsumura and Ananthaswamy, 2004).

There is also strong evidence suggesting that DNA damage itself contributes to immune suppression. For example, Kripke et al. (1992) showed that repairing CPDs in UVB irradiated mice using the T4N5 excision repair enzyme prevented UV-induced suppression of CHS and DHS. In addition, a study by Nishigori et al. (1996) demonstrated that keratinocytes pretreated with T4N5 to repair UV-induced lesions did not produce the immunosuppressive cytokine IL-10. This confirms that UV radiation suppresses the immune system not only directly, but also indirectly through photoproduct formation. While this mechanism may have evolved to prevent autoimmune responses against the inflammatory byproducts of UV damage, the downside is that it harms human health by driving carcinogenesis (Rabe et al., 2006). This is reflected in the

increased skin cancer risk demonstrated by patients undergoing immunosuppressive therapies (Matsumura and Ananthaswamy, 2004).

3. Effect of UV radiation on the clinical level

3.1. Sunburn and tanning

The main clinical consequences of UV exposure includes sunburn and tanning, both of which are influenced by genetic predispositions, primarily skin pigmentation (Matsumura and Ananthaswamy, 2004). A classification system called Fitzpatrick SPT is used to classify the tendency to burn or tan based on skin color (individuals with very light skin tend to burn more and tan less and vice versa) (Fig. 6).

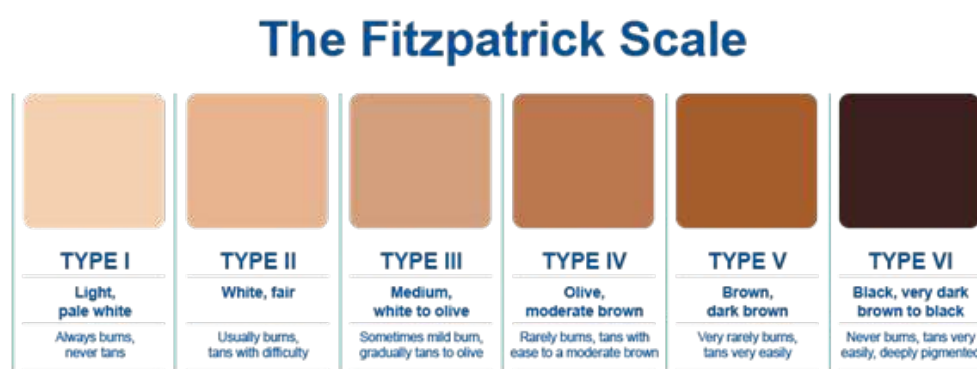


Figure 6. Fitzpatrick SPT scale (Sutton)

Sunburn is an acute inflammatory response that involves the vasodilation of blood vessels in the epidermis. UV-B radiation plays the major role in inducing sunburn erythema, being 1000-fold more efficient than long-wave UV-A radiation (Baron and Suggs, 2014). Upon UV exposure, the transcription factor (NF)- κ B is activated, which in turn drives the production of proinflammatory cytokines, including IL-1, IL-6, TNF- α , and VEGF. These cytokines attract neutrophils (a type of white blood cell) to the damaged site, and produce free radicals that cause oxidative DNA damage (Rabe et al., 2006).

Tanning, on the other hand, is a protective response which occurs in two phases: immediate pigment darkening (IPD) and delayed tanning (DT). IPD happens within seconds after UV exposure due to redistribution of melanosomes (organelles that contain melanin) and oxidation of melanin (Matsumura and Ananthaswamy, 2004; Baron and Suggs, 2014). It occurs in individuals with skin types III-IV on the Fitzpatrick SPT scale (Rabe et al., 2006). It is suggested that IPD protects epidermal basal cells from damage by forming nuclear caps (Soter, 1990). DT, the second phase, occurs 2-3 days after exposure and it is a result of both UV-A and UV-B radiation. UV-B-induced DT primarily involves an increase in activity and number of melanocytes. This phase also involves an increase in melanin production (melanogenesis) and melanin transfer to keratinocytes, causing more skin pigmentation (Matsumura and

Ananthaswamy, 2004). A critical player in the process of melanogenesis is melanocortin 1 receptor (MC1R), which is activated by UV-triggered hormones to drive melanogenesis (Baron and Suggs, 2014). In the case of UV-A-induced DT, longer wavelengths (340-400 nm) cause increased melanin density, while shorter wavelengths (320-340) cause increased melanin production and transfer (Matsumura and Ananthaswamy, 2004). This increased pigmentation creates protection from further UV-induced damage, as melanin absorbs and scatters UV, decreasing its penetration through the skin. However, melanin does not provide full protection, only absorbing around 50-75% of UV and having an SPF of around 1.5-2 (Baron and Suggs, 2014). By comparison, most sunscreens have an SPF of 30 or above. As a result, UV-induced lesions can continue to accumulate despite tanning.

Besides sunburn and pigmentation, the skin also reacts to UV radiation through hyperplasia or thickening of the stratum corneum, epidermis, and dermis. Around 1-2 days after UV exposure, following the activation of p53, mitotic activity increases in the epidermis and dermis, providing protection from further UV damage (Matsumura and Ananthaswamy, 2004). The photoprotective properties of this process have been shown in patients with vitiligo, who due to their lack of melanin in certain areas, rely on hyperplasia for protection (EVERETT, 1961).

Conclusion

UV radiation induces molecular, immunological, and cellular damage to human skin through a multitude of intertwined mechanisms. UV-A, the major component of UV, causes oxidative stress to the DNA of epidermal cells, while UV-B, the most carcinogenic component, causes the formation of photoproducts which disrupt DNA's structure. DNA repair pathways and cellular stress responses are employed to resolve the damage, however, if these prove unsuccessful, mutations could occur in specific tumor-related genes, resulting in the development of NMSC or melanoma. Contributing to the development of skin cancer is also the UV-induced weakening of the immune system and the cellular response of erythema (sunburning). Taken together, the evidence clarifies the links between UV radiation and skin cancer, establishing UV as a powerful carcinogen. From a public health perspective, the findings underscore the need to improve UV protection strategies, such as building awareness campaigns, strengthening sunscreen recommendations by dermatologists, and regulating the use of artificial UV sources. From a scientific perspective, the paper highlights the frequently overlooked effects of UV-A radiation, and demonstrates the overlap between UV-A and UV-B, which are frequently portrayed as distinct in their effect and mechanisms. This broader perspective could contribute to the development of successful therapeutic approaches and improved sunscreen formulations.

At the same time, the specific pathways of melanoma development and the exact mechanism of UV-A (including the identity of its chromophores and its contribution to each type of skin cancer) remain unclear, due to limited and often conflicting evidence. As a result, further research is needed for scientists to fully understand the role of UV exposure on skin cancer development, thus paving the way for more effective prevention strategies and medical treatments.

Works Cited

- Baron, Elma D., and Amanda K. Suggs. "Introduction to Photobiology." *Dermatologic Clinics*, vol. 32, no. 3, July 2014, pp. 255–66, <https://doi.org/10.1016/j.det.2014.03.002>.
- Besaratinia, Ahmad, et al. "Riboflavin Activated by Ultraviolet A1 Irradiation Induces Oxidative DNA Damage-Mediated Mutations Inhibited by Vitamin C." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 14, National Academy of Sciences, Apr. 2007, pp. 5953–58, <https://doi.org/10.1073/pnas.0610534104>.
- Cadet, Jean, et al. "Characterization of Thymidine Ultraviolet Photoproducts. Cyclobutane Dimers and 5,6-Dihydrothymidines." *Canadian Journal of Chemistry*, vol. 63, no. 11, NRC Research Press, Nov. 1985, pp. 2861–68, <https://doi.org/10.1139/v85-477>.
- Caldecott, Keith W. "DNA Single-Strand Break Repair and Human Genetic Disease." *Trends in Cell Biology*, May 2022, <https://doi.org/10.1016/j.tcb.2022.04.010>.
- Cancer Research UK. "How Does the Sun and UV Cause Cancer?" Cancer Research UK, CRUK, 30 Nov. 2023, www.cancerresearchuk.org/sites/default/files/styles/cruk_wide_resp_breakpoint_two/public/types_of_uv_rays_higher_res.png?itok=nV4fOPa4.
- Carter, Rachel J., and Jason L. Parsons. "Base Excision Repair, a Pathway Regulated by Posttranslational Modifications." *Molecular and Cellular Biology*, vol. 36, no. 10, May 2016, pp. 1426–37, <https://doi.org/10.1128/MCB.00030-16>.
- Cleveland Clinic. "Lymph Node Function & Location." Cleveland Clinic, 31 May 2022, my.clevelandclinic.org/health/body/23131-lymph-nodes.
- Cleveland Clinic. "Skin Cancer Overview | Cleveland Clinic." Cleveland Clinic, 19 Nov. 2021, my.clevelandclinic.org/health/diseases/15818-skin-cancer.
- Douki, Thierry, and Jean Cadet. "Individual Determination of the Yield of the Main UV-Induced Dimeric Pyrimidine Photoproducts in DNA Suggests a High Mutagenicity of CC Photolesions." *Biochemistry*, vol. 40, no. 8, Feb. 2001, pp. 2495–501, <https://doi.org/10.1021/bi0022543>.
- EVERETT, MARK ALLEN. "Protection from Sunlight in Vitiligo." *Archives of Dermatology*, vol. 84, no. 6, American Medical Association (AMA), Dec. 1961, p. 997, <https://doi.org/10.1001/archderm.1961.01580180113017>.
- Hall, Daniel B., et al. "Oxidative DNA Damage through Long-Range Electron Transfer." *Nature*, vol. 382, no. 6593, Nature Portfolio, Aug. 1996, pp. 731–35, <https://doi.org/10.1038/382731a0>.
- Center for Devices and Radiological Health. "Ultraviolet (UV) Radiation." FDA, Aug. 2020, www.fda.gov/radiation-emitting-products/tanning/ultraviolet-uv-radiation#UVC.
- Hernández Borrero, Liz J., and Wafik S. El-Deiry. "Tumor Suppressor P53: Biology, Signaling Pathways, and Therapeutic Targeting." *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1876, no. 1, Aug. 2021, p. 188556, <https://doi.org/10.1016/j.bbcan.2021.188556>.
- Hill, Laurie, et al. "Fas Ligand: A Sensor for DNA Damage Critical in Skin Cancer Etiology." *Science*, no. 5429, Aug. 1999, pp. 898–900, <https://doi.org/10.1126/science.285.5429.898>.
- Hindi, Nagham Nafiz, et al. "The Base Excision Repair Process: Comparison between Higher

- and Lower Eukaryotes.” *Cellular and Molecular Life Sciences: CMLS*, vol. 78, no. 24, Dec. 2021, pp. 7943–65, media.springernature.com/lw1200/springer-static/image/art%3A10.1007%2Fs00018-021-03990-9/MediaObjects/18_2021_3990_Fig3_HTML.png.
- Iordanou, Eleni, and Mark Berneburg. “Phototherapy and Photochemotherapy.” *JDDG Journal Der Deutschen Dermatologischen Gesellschaft*, vol. 8, no. 7, Wiley, June 2010, pp. 533–40, <https://doi.org/10.1111/j.1610-0387.2010.07442.x>.
- Ito. “8-Hydroxydeoxyguanosine Formation at the 5’ Site of 5’-GG-3’ Sequences in Double-Stranded DNA by UV Radiation with Riboflavin.” *The Journal of Biological Chemistry*, vol. 268, no. 18, J Biol Chem, 2025, pubmed.ncbi.nlm.nih.gov/8390459.
- Jiang, Jun, et al. “Polymorphisms of DNA Repair Genes: ADPRT, XRCC1, and XPD and Cancer Risk in Genetic Epidemiology.” *Methods in Molecular Biology (Clifton, N.J.)*, vol. 471, 2009, pp. 305–33, https://doi.org/10.1007/978-1-59745-416-2_16.
- John, Juliana. “Fenton Reaction - an Overview | ScienceDirect Topics.” *Www.sciencedirect.com*, 2022, www.sciencedirect.com/topics/earth-and-planetary-sciences/fenton-reaction.
- Kimura, Seisuke, and Kengo Sakaguchi. “DNA Repair in Plants.” *Chemical Reviews*, vol. 106, no. 2, Jan. 2006, pp. 753–66, www.researchgate.net/publication/7310454/figure/fig3/AS:277719284764683@1443224845147/Nucleotide-excision-repair-NER-NER-pathway-s-can-be-classified-into-two-groups-global.png.
- Kripke, M. L., et al. “Pyrimidine Dimers in DNA Initiate Systemic Immunosuppression in UV-Irradiated Mice.” *Proceedings of the National Academy of Sciences*, vol. 89, no. 16, Aug. 1992, pp. 7516–20, <https://doi.org/10.1073/pnas.89.16.7516>.
- Krokan, H. E., and M. Bjoras. “Base Excision Repair.” *Cold Spring Harbor Perspectives in Biology*, vol. 5, no. 4, Apr. 2013, pp. a012583–83, <https://doi.org/10.1101/cshperspect.a012583>.
- L.A. Burns-Naas. “Hypersensitivity, Delayed Type.” Elsevier EBooks, Elsevier BV, Jan. 2014, pp. 986–88, <https://doi.org/10.1016/b978-0-12-386454-3.00032-4>.
- Langselius, Oliver, et al. “Global Burden of Cutaneous Melanoma Incidence Attributable to Ultraviolet Radiation in 2022.” *International Journal of Cancer*, Wiley, May 2025, <https://doi.org/10.1002/ijc.35463>.
- Matsumura, Yasuhiro, and Honnavara N. Ananthaswamy. “Toxic Effects of Ultraviolet Radiation on the Skin.” *Toxicology and Applied Pharmacology*, vol. 195, no. 3, Mar. 2004, pp. 298–308, <https://doi.org/10.1016/j.taap.2003.08.019>.
- Mullenders, Leon H. F. “Solar UV Damage to Cellular DNA: From Mechanisms to Biological Effects.” *Photochemical & Photobiological Sciences*, vol. 17, no. 12, Dec. 2018, pp. 1842–52, <https://doi.org/10.1039/C8PP00182K>.
- Nishigori, C., et al. “Evidence That DNA Damage Triggers Interleukin 10 Cytokine Production in UV-Irradiated Murine Keratinocytes.” *Proceedings of the National Academy of Sciences*, vol. 93, no. 19, Sept. 1996, pp. 10354–59, <https://doi.org/10.1073/pnas.93.19.10354>.
- Nishigori, C., et al. “Mutations in Ras Genes in Cells Cultured from Mouse Skin Tumors

- Induced by Ultraviolet Irradiation.” *Proceedings of the National Academy of Sciences*, vol. 91, no. 15, National Academy of Sciences, July 1994, pp. 7189–93, <https://doi.org/10.1073/pnas.91.15.7189>.
- Pouget, J. -P, et al. “DNA Damage Induced in Cells by γ and UVA Radiation as Measured by HPLC/GC–MS and HPLC–EC and Comet Assay.” *Chemical Research in Toxicology*, vol. 13, no. 7, American Chemical Society, June 2000, pp. 541–49, <https://doi.org/10.1021/tx000020e>
- Rabe, Jessica H., et al. “Photoaging: Mechanisms and Repair.” *Journal of the American Academy of Dermatology*, vol. 55, no. 1, July 2006, pp. 1–19, <https://doi.org/10.1016/j.jaad.2005.05.010>.
- Ravanat, Jean-Luc, et al. “Direct and Indirect Effects of UV Radiation on DNA and Its Components.” *Journal of Photochemistry and Photobiology B: Biology*, vol. 63, no. 1-3, Oct. 2001, pp. 88–102, [https://doi.org/10.1016/s1011-1344\(01\)00206-8](https://doi.org/10.1016/s1011-1344(01)00206-8).
- Reardon, Joyce T., and Aziz Sancar. “Nucleotide Excision Repair.” *Progress in Nucleic Acid Research and Molecular Biology*, vol. 79, 2005, pp. 183–235, [https://doi.org/10.1016/s0079-6603\(04\)79004-2](https://doi.org/10.1016/s0079-6603(04)79004-2).
- Reid, and Lawrence A. Loeb. “Tandem Double CC-->TT Mutations Are Produced by Reactive Oxygen Species.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 9, National Academy of Sciences, May 1993, pp. 3904–7, <https://doi.org/10.1073/pnas.90.9.3904>.
- Ridley, Andrew J., et al. “Cellular and Sub-Cellular Responses to UVA in Relation to Carcinogenesis.” *International Journal of Radiation Biology*, vol. 85, no. 3, Jan. 2009, pp. 177–95, <https://doi.org/10.1080/09553000902740150>.
- Roky, Amdad Hossain, et al. “Overview of Skin Cancer Types and Prevalence Rates across Continents.” *Cancer Pathogenesis and Therapy*, vol. 3, no. 2, Elsevier BV, Aug. 2024, <https://doi.org/10.1016/j.cpt.2024.08.002>.
- Rünger, T. M. “Role of UVA in the Pathogenesis of Melanoma and Non-Melanoma Skin Cancer: A Short Review.” *Photodermatology, Photoimmunology & Photomedicine*, vol. 15, no. 6, Dec. 1999, pp. 212–16, <https://doi.org/10.1111/j.1600-0781.1999.tb00090.x>.
- Setlow, R. B., et al. “Wavelengths Effective in Induction of Malignant Melanoma.” *Proceedings of the National Academy of Sciences*, vol. 90, no. 14, July 1993, pp. 6666–70, <https://doi.org/10.1073/pnas.90.14.6666>.
- Soter, N. A. “Acute Effects of Ultraviolet Radiation on the Skin.” *Seminars in Dermatology*, vol. 9, no. 1, Mar. 1990, pp. 11–15, pubmed.ncbi.nlm.nih.gov/2203437/.
- Soufir, Nadem, et al. “P16 UV Mutations in Human Skin Epithelial Tumors.” *Oncogene*, vol. 18, no. 39, Sept. 1999, pp. 5477–81, <https://doi.org/10.1038/sj.onc.1202915>.
- Starcevic. “Is There a Link between DNA Polymerase Beta and Cancer?” *Cell Cycle* (Georgetown, Tex.), vol. 3, no. 8, Cell Cycle, 2017, pubmed.ncbi.nlm.nih.gov/15280658/.
- Sutton. “What’s Your Type? - Sutton Dermatology + Aesthetics.” *Suttonderm.com*, 17 Oct. 2016, suttonderm.com/content/uploads/2016/09/FITZPATRICK-COLOR-CHART.png.
- Svobodova, Alena, et al. “ULTRAVIOLET LIGHT INDUCED ALTERATION to the SKIN.” *Biomedical Papers*, vol. 150, no. 1, July 2006, pp. 25–38, www.researchgate.net/publication/6851522/figure/fig1/AS:340527720484864@1458199

- 543019/Formation-of-pyrimidine-dimer-from-adjacent-pyrimidine-bases-on-the-same-s and-A.png.
- Tornaletti, Silvia, and Gerd P. Pfeifer. "UV Damage and Repair Mechanisms in Mammalian Cells." *BioEssays*, vol. 18, no. 3, Mar. 1996, pp. 221–28, <https://doi.org/10.1002/bies.950180309>.
- WHO. "Skin Cancer – IARC." [Www.iarc.who.int](http://www.iarc.who.int), 2022, www.iarc.who.int/cancer-type/skin-cancer/.
- Wu, Shaowei, et al. "History of Severe Sunburn and Risk of Skin Cancer among Women and Men in 2 Prospective Cohort Studies." *American Journal of Epidemiology*, vol. 183, no. 9, Apr. 2016, pp. 824–33, <https://doi.org/10.1093/aje/kwv282>.
- Yang, Haozhe, et al. "Synthesis of Oligonucleotides Containing the N-(2-Deoxy- α,β -D-Erythropentofuranosyl)-2,6-Diamino-4-Hydroxy-5-Formamidopyrimidine (Fapy·DG) Oxidative Damage Product Derived from 2'-Deoxyguanosine." *Chemistry - a European Journal*, vol. 26, no. 24, Wiley, Apr. 2020, www.researchgate.net/publication/340542054/figure/fig1/AS:11431281245958208@1716261982306/Formation-of-FapydG-and-8-OxodGuo.png.

Descriptive Analysis of NBA Shooting Trends: A Comparative Study of Five Teams in the Most Recent Regular Season By Adish Reddy and Neharika Shrivastava

Abstract

This study is a comparative statistical analysis of five NBA teams selected from the official NBA website in an alphabetical order on different performance metrics like field goal percentage, free throw percentage, assists, personal fouls, disqualification, steals, turnover and blocks. The study revealed that most performance metrics remained consistent across teams, free throw efficiency and shot-blocking ability varied significantly, affecting overall team performance. The study recommended that teams with lower FT% can focus on free throw drills, biomechanical analysis and can enhance mental conditioning of the players. Teams with lower BLK like Chicago Bulls can focus on rim protection training and positional awareness drills for tall men. Turnovers did not show significant differences, but teams like Charlotte Hornet can prioritize ball security through structured offensive play, decision making training and passing accuracy drills. This study is limited to only descriptive analytics of 5 NBA teams for the past 8 years only for the regular season. Future study can include more advanced data analysis techniques for predicting players and team performance based on individual shooting mechanics, fatigue levels, and game conditions

Keywords

NBA, basketball, FG%, FT%, BLK, STL, TO, descriptive statistics, average

Introduction

In basketball for evaluating players performance and framing team strategies, evaluating their past performance using statistical data to gain deeper insights has evolved significantly. Some of the recent research has emphasized the growing importance of analyzing the past and present data to predict the performance ((Sarlis and Tjortjis), (Pan et al.)).

Researchers have examined that for National Basketball Association (NBA) players that socio- economic, demographic and injury related factors impact the players performance. As per the finding peak performance generally occurs between the age of 27 and 29 and players earn the highest salary between the ages of 29 and 34. Salary structure and career longevity is significantly influenced by players position and age. Musculoskeletal injuries account for nearly 50% of financial losses due to health-related issues in the NBA (V. Sarlis and Tjortjis). It was also analysed that sports psychology plays a crucial role in enhancing athletic performance by helping professional basketball players manage challenges, maintain focus on their goals, and recover from injuries. According to a regression analysis there is a strong positive relationship between administrative support and athletic performance, with psychological capital serving as a key mediating factor that enhances the impact of organizational support on player success (Pang et al.) According to a study higher shooting accuracy (True Shooting Percentage, TS%) and

strong defensive metrics (Defensive Rating, DRtg) are crucial predictors of success in NCAA Division I collegiate basketball and the NBA. Whereas in the NBA, Player Efficiency Rating (PER) plays a more significant role in determining team performance (Liang et al.). In fact for team success effective field goal percentage (FG%), turnover percentage (TOV%), offensive rebound percentage (ORB%), and free throw rate (FTR) have been identified as key indicators (Kotzias). A machine learning model was trained using the historical data from 2013/2014 to 2017/2018 NBA seasons incorporating an Optimal Time Window (OTW) to select the most relevant training data. The model achieved an average prediction accuracy of 66%, with a maximum accuracy of 78%, by analysing basic, advanced, and league-wide metrics (Horvat et al.)

This study also aims to analyze the performance of NBA teams by comparing the descriptive statistics of various performance metrics over the past seven seasons. Team data was collected from the official NBA website, focusing on key statistical indicators such as field goals made, total assists, and other relevant performance ratios. A comprehensive analysis was conducted using descriptive statistics to identify significant differences in team performance trends over time. These insights contribute to a deeper understanding of how teams evolve and adapt their strategies for sustained success.

2. Literature Analysis

2.1 Data Analytics in Basketball

Performance analytics in Basketball incorporating advanced statistics has evolved significantly to analyse player performance, team strategies and decision making (Gutiérrez). A predictive modelling was designed using Naïve Bayesian classification to forecast game outcomes based on key metrics like field goal percentage, turnovers, free throws, and rebounds (M. Libed et al.). By using a machine learning tool called Classification and Regression Trees (CART) researchers developed court maps to visualize scoring probability. This model provides better spatial accuracy, helping teams refine offensive and defensive strategies (Zuccolotto et al.). Using some more advanced basketball analytics it was found out that offensive round predictions reached 67.94% accuracy, while specific player rotations optimized scoring efficiency ((Zheng et al.), (Fei Lu and Hao Li)). Similarly, NBA game outcomes were also predicted by researchers by applying machine learning techniques using three seasons' data (2020/21–2022/23). According to the findings, A non-linear variable selection algorithm identified the most impactful half-time statistics while maintaining high predictive performance (Tsagris et al.). In fact, a researcher also developed a model using deep learning techniques which helps in posture recognition, aiding technical movement correction, muscle memory development, and performance optimization (Yao).

2.2 Physical and Biomechanical Aspects of Basketball Performance

Physical fitness has a high correlation with basketball performance. Amongst Spanish 2nd Division professional players data revealed a progressive decline in physical performance across game quarters, with total distance covered and player load experiencing the most

significant reductions (García et al.). It also examined the relationship between strength, flexibility, and balance in elite male basketball players and it was found out that there is a strong correlation between dynamic balance and performance (Özer). According to a biochemical study lower limb movement impacts shooting efficiency. It was found out that force plates and motion tracking showed that jump shots exhibited superior biomechanical characteristics which optimizes power output and shot stability (Struzik et al.). In fact, amongst adolescent female players, shooting adjustments at different distances involved increasing ball velocity while decreasing release angle, suggesting necessary training adaptations (França et al.).

2.3 Training Methodologies and Their Impact on Player Development

Research in basketball has significantly increased especially related to maturation, athletic performance, and training load. Study suggests that training intensity and player maturity strongly influence performance outcomes (SOFYAN et al.). High intensity interval training (HIIT) is found to be an effective conditioning tool which significantly improves cardiovascular endurance, agility, and power. According to a meta-analysis it was suggested that endurance tests and moderate effects on sprint performance, suggesting that HIIT should be strategically integrated into training regimens (Cao et al.). Even TRX suspension training has proven beneficial in enhancing strength, stability, and aerobic capacity, making it a valuable addition for conditioning programs (Tyshchenko et al.).

2.4 Socioeconomic and Psychological Factors in Basketball

Social and psychological factors influence basketball participation and performance. A study revealed using gender-based analysis using Eurobarometer data revealed that females participate in sports at lower rates than males, influenced by factors such as education, employment, and geographical location. It has been suggested that policymakers are encouraged to implement tailored strategies to boost female participation in sports (Öztürk et al.). Interestingly, the Chinese Basketball Association (CBA) has seen economic growth due to foreign player inclusion, leading to increased ticket sales and sponsorships. However there are concerns regarding imbalanced player development and reduced advertising value for domestic talent suggest the need for improved talent management strategies (Liu). Mental attributes like confidence, attention, and emotional stability significantly impact shooting accuracy. Training methods incorporating psychological resilience exercises have been recommended to enhance player focus and shooting efficiency (Tsiannis and Efstathiades). An assessment was done to analyze the impact of different basketball exercise intensities on college students' mental well-being. It was found out that it has a positive effect on stress relief and psychological health (Zhang).

Some key performance indicators (KPI) in basketball have evolved due to integration of data science in sports. A study was conducted to compare the National Collegiate Athletic Association (NCAA) Division I and NBA metrics and it was found that in collegiate basketball, shooting accuracy and defensive metrics were the strongest predictors of success, while in the NBA, Player Efficiency Rating (PER) played a more significant role (Liang et al.).

Strong correlation was found in sitting height and basketball performance, with moderate to strong correlations observed between sitting height and scoring, rebounding, and shot-blocking abilities (Shamim).

Hence as basketball analytics continues to evolve, integration of data analytics advanced biomechanical research is expected to further enhance training methodologies, performance evaluation, and game strategy development. This study aims to contribute to the literature by identifying and comparing the significant ratios impacting the NBA team performance over a time period of 7 years.

3. Research Methodology

For the study, the data was collected for the top five NBA teams arranged in alphabetical order based for their regular-season performance over the past seven years from the official NBA website (*Cume Stats | Stats | NBA.com*). For each team yearly performance was recorded in Microsoft Excel. The selected teams for the study are ‘Atlanta Hawks’, ‘Boston Celtics’, ‘Brooklyn Nets’, ‘Charlotte Hornets’ and ‘Chicago Bulls’

3.1 Key Performance Metrics Analyze

The study examined various statistical measures which are crucial for evaluating a team's performance and efficiency in the NBA. The selected key ratios are field goal percentage (FG%) which measures the team shooting efficiency. It is the ratio of the number of successful field goals made divided to the total field goals attempted. It counts all the shots taken during live game action; it does not include free throws which are measured separately as free throw percentage. Higher FG% indicates better shooting accuracy and offensive efficiency (Raptors). Another important metric is free throw percentage (FT%) which measures the accuracy of a team's free throw shooting which is calculated as the number of successful free throws made divided by total free throw attempts. It evaluates the players ability to capitalize the scoring opportunity which reflects both their skills and confidence at the free throw line. Hence, strong (FT%) is essential for maximizing scoring opportunities from fouls (“Free Throw Percentage - (Sports Journalism) - Vocab, Definition, Explanations | Fiveable”).

Assist (AST) records the number of passes which lead to a successful field goal. High assist numbers reflect strong teamwork, ball movement, and offensive coordination (Supola et al.). Personal Fouls (PF) represents the tracks number of fouls committed by a team during gameplay and excessive personal fouls can lead to free throw opportunities for the opposing team which can impact the game outcomes and the number of players who have fouled out during the game is called disqualifications (DQ) (Kotzias, “The Four Factors of Basketball as a Measure of Success”). High DQ numbers may indicate aggressive defensive play or poor foul discipline. Possession of the ball from an opponent team without committing foul is an important metric in any basketball game which is measured by a metric called steals (STL). A high number of steals suggests strong defensive pressure and disruption of the opponent's offense and the number of times a team loses possession due to mistakes such as bad passes, traveling, or offensive fouls is represented by turnover (TO) (Kotzias, “The Four Factors of Basketball as a

Measure of Success”). Lower turnovers indicate better ball control and decision-making. In a basketball game Blocks (BLK) measures the number of times a team successfully deflects an opponent’s shot attempt. Strong shot-blocking ability can indicate a team’s defensive strength in protecting the basket (Pratama).

By analysing these performance indicators over eight years for regular seasons, the study aims to identify trends and patterns that contribute to team success in the NBA. This descriptive analytical approach provides valuable insights into the strengths and weaknesses of each selected team.

4. Data Analysis

The collected data from the NBA website was arranged in a time series format for each selected team over a period of 8 years. The collected data for each team has been depicted in bar diagrams. The fig 1 depicts the performance of the Atlanta Hawk Team.

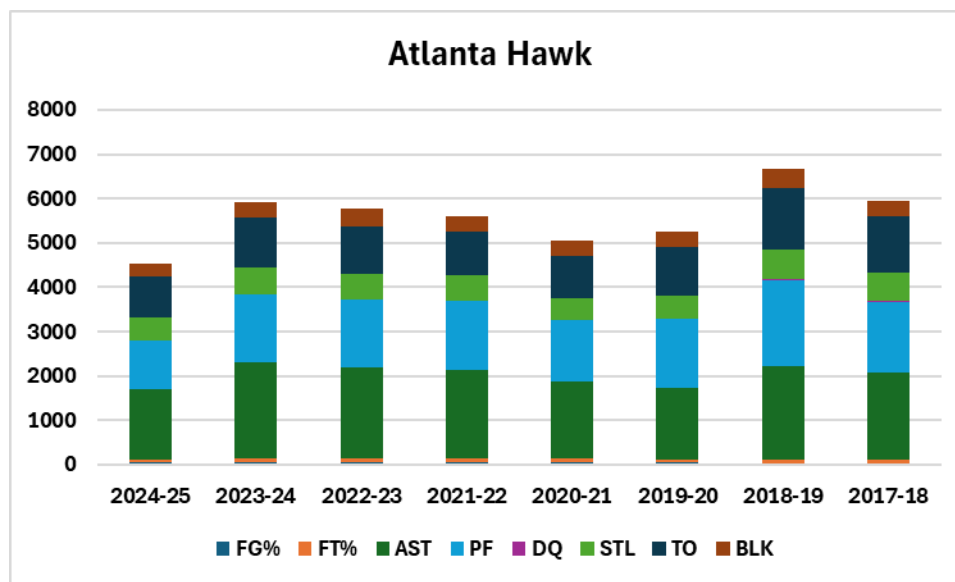


Fig 1. Atlanta Hawk Team

As per the data of Atlanta Hawk, the team has grown steadily in offensive efficiency (FG%) and defensive discipline (DQ, TO). 2018 was the most volatile season for the team with high turnovers, fouls and disqualification, but post 2020 the team has become more efficient, less foul prone and more accurate from the field. The team has shown continuous improvement in turnovers and FG% which reflects matured gameplay and strategy

Below Fig 2 represents the data of the team Boston Celtic

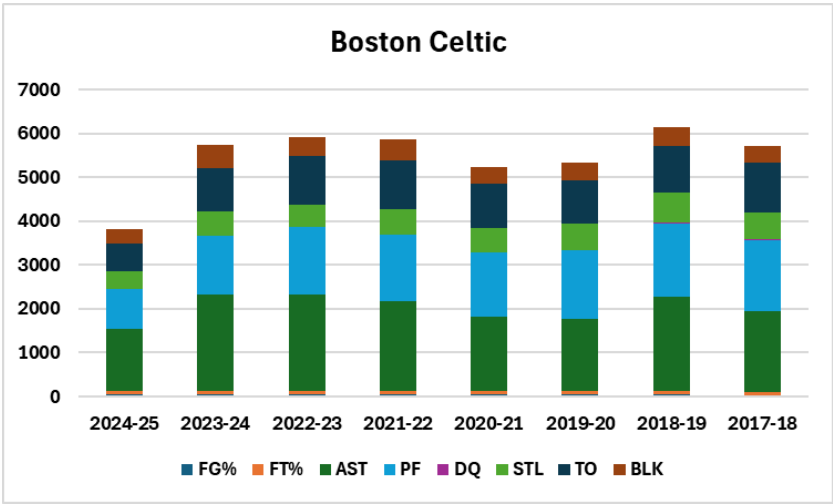


Fig 2. Boston Celtic Team

For Boston Celtic in 2017-18 was a rebuilding year, as they had lowest FG%, FT% and high TO. From 2018-19 till 2023-24 for the team it was a period of high assist, increased FG% and strong defensive stats. In 2024-25 the team experienced a drop in AST and PF due to low paced and more isolation-based offense, but they also showed the sign of maturity and disciplined gameplay by best turnover numbers and fewest fouls. They also changed their defensive scheme, hence blocking and stealing dropped a bit. Fig 3 below represents the chart of Brooklyn Net

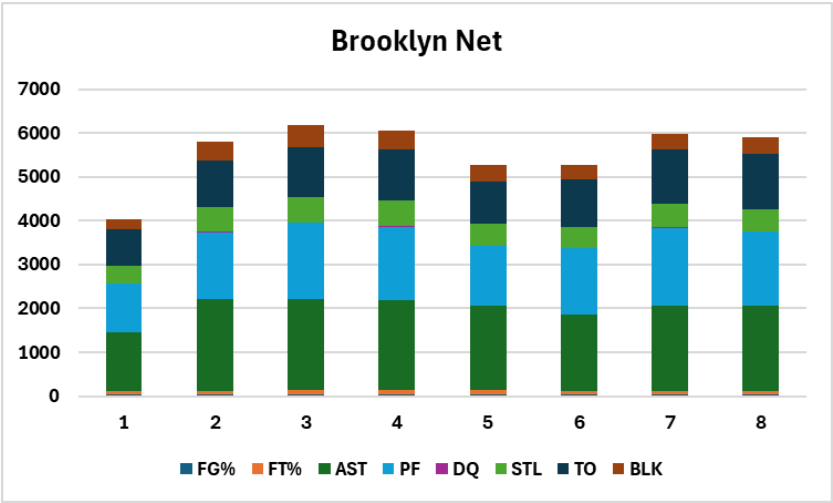


Fig 3. Brooklyn Net

The descriptive statistics of each selected variable for each team has been described in table 1.

The team experienced notable fluctuation in the performance which is reflecting shifts in roster dynamic and strategic focus. In the 2020-21 season the team saw a peak of 49.4% of FG% indicating a highly efficient offensive year, while 2019-20 marked a low point at 34.3%. Free throw accuracy remained relatively strong, with consistent improvement post 2020, culminating in an 80.5% high in 2020-21. Team coordination as measured by assists, reached its zenith in 2023-24 with 2102 assists, but dropped significantly next year in 2024-25. The team showed discipline with reduction in fouls and disqualification declining to lowest in 2024-25. In the same year the team also experienced a sharp dip in blocks and steals. Overall, the team has shown resilience and adaptability, even in the most recent seasons. The below fig 4 represents the data of Charlotte Hornet team

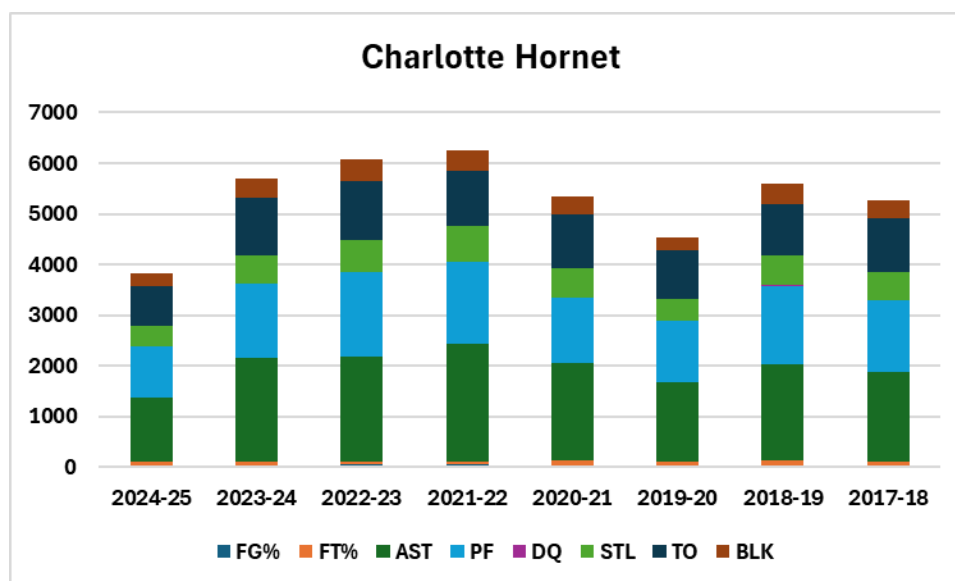


Fig 4. Charlotte Hornet

The Charlotte Hornets' team performance reflects that team has evolved its offensive strategies and defensive challenges. The team average field goal % stood at 43.7%, with a minimum in the year 2023-24, indicating offensive struggle that season. FT% remained steady with an average of 76.3% and maximum at 79.7% in the year 2018-19. The team displayed a strong playmaking ability, with assists reaching a high of 2302 in 2021-22 which indicates peak in ball movement and offensive coordination. On the defensive end, the team averaged 556.9 steals and 353.8 blocks per person and in the year 2022-23 the team showed the highest block count at 425. In spite of these highlights, turnover averaged around 1000 per season, which is revealing consistency in possession losses. Disqualification also increased in 2028-19 to 16, which raises questions about on – court discipline

Below fig 5 exhibits the bar diagram for the team Chicago Bull

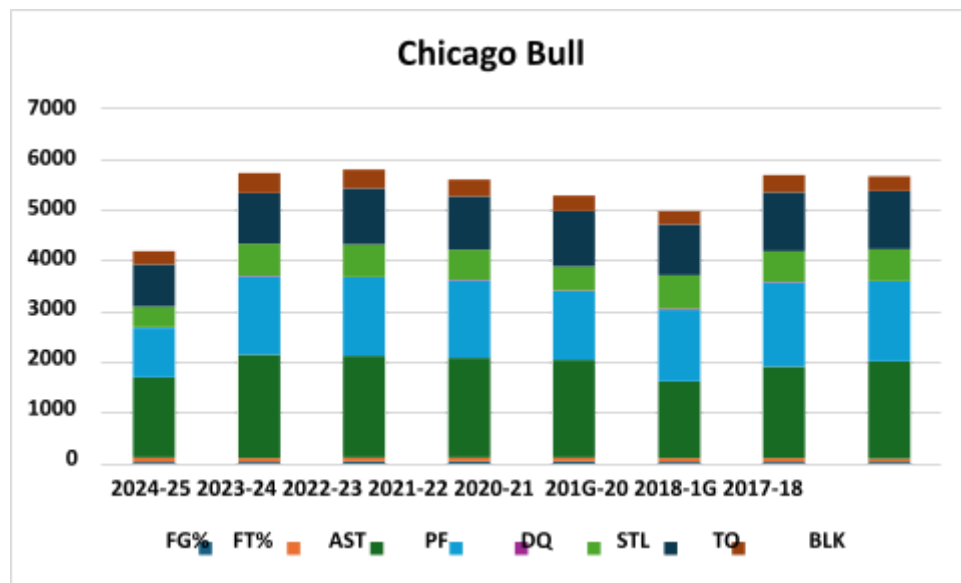


Fig 5. Chicago Bull

The Chicago Bull team have demonstrated a strong offensive profile in the last 8 years of regular season, with an average FG% of 44%, maximum of 49% in the year 2022-23. Their free throw is consistent, with an average of 78.1% and best performance in the year 2021-22 at 81.3%. In terms of playmaking, the team recorded an average of 1837 assists per person, with a standout in the year 2023-24 which reflects effective ball distribution. In the defensive game, the team averages 575.5 steals and 329.9 blocks per person, with high impact defensive performance in the year 2022-23. However, the turnovers remained consistent indication areas for better decision making and ball control. Disqualification spiked in the year 2019-20 pointing to a rare lapse in court discipline. The overall team maintained competitive numbers across most metrics with notable strength in shooting efficiency and assists.

To compare the team performance for the past 8 years regular season game descriptive statistics was calculated using MS-Excel. Table 1 depicts the descriptive statistics of all the five teams for different variables.

Table 1: Descriptive Statistics of NBA Selected Teams

Team	Statistics	FG%	FT%	AST	PF	DQ	STL	TO	BLK
Atlanta Hawk	Mean	43.8	79.22	1902.2	1522.37	9.37	580.12	1095.6	356.6
	Median	46.4	79.35	1981.5	1537.5	8.5	583.5	1073	348
	Standard Deviation	5.19	2.25	235.7	229.27	4.47	61.45	166.12	40.9
Boston Celtic	Mean	45.71	79.71	1898.75	1450.12	4.75	566.37	1006	419.62
	Median	46.55	80.15	1939	1531.5	3.5	574	1032	418
	Standard Deviation	3.35	1.64	292.24	244.6	3.15	86.13	156.33	68.26

Brookly Net	Mean	44.8	77.77	1898	1543.25	9.375	516.62	1093	379.87
	Median	45.25	78.35	1947	1593	8.5	525.5	1113	384.5
	Standard Deviation	4.71	2.63	254.5	223.5	4.34	60.9	136.35	84.48
Charlotte Hornet	Mean	43.7	76.33	1849.5	1406.5	8.25	556.8	1029.12	353.75
	Median	44.9	75.5	1919	1440.5	8	563.5	1055	372
	Standard Deviation	3.53	2.12	329.35	220.9	4.89	98.72	117.23	66.15
Chicago Bull	Mean	44.03	78.73	1843.87	1453.5	7.75	580.12	1047.75	321.5
	Median	45.85	79.1	1925	1540.5	5.5	614.5	1071	320
	Standard Deviation	5.36	2.11	198.55	211.59	6.4	87.40	105.83	49.10

As per the descriptive statistics, the Atlanta Hawk team has a relatively low FG% (Mean: 43.8%), which indicates struggles in scoring efficiently compared to other teams, whereas their FT% (79.22%) is strong, showing they capitalize well on free throws. Atlanta averages 1902.2 assists per season, showing decent ball distribution, but they are not among the best passing teams, but with 580.12 steals per season, the Hawks demonstrate an above-average defensive presence. The team commits 1095.6 turnovers on average, suggesting room for improvement in ball control. Additionally, high personal fouls (1522.37) indicate a tendency to play aggressive defense. The Hawks record a relatively high number of disqualifications (9.37 per season) and average 356.6 blocks per person, indicating a below-average rim-protection ability. Over all Hawks struggle with field goal accuracy but maintain a strong defensive presence with blocks and steals. High turnovers and fouls remain areas for improvement for Atlanta Hawk. Boston Celtic has a high FG% (45.71%) and the highest FT% (79.71%), making them one of the more

efficient teams offensively. With 1898.75 assists per season, the Celtics demonstrate strong ball movement but slightly less than teams like Brooklyn. The Celtics average 566.37 steals per season, showing a respectable defensive effort. Boston averages 1006 turnovers per season (the lowest among the teams), showing strong ball control. Their personal foul count (1450.12) is also moderate. The team records 419.62 blocks per season, which is among the lowest in this study, suggesting the team should improve in rim protection. The Boston Celtics excel in free throw shooting, turnovers control, and team play. They are one of the most disciplined teams, but their blocking ability is lower than other teams. Brooklyn Net has a moderate FG% (44.8%) and relatively lower FT% (77.77%), indicating average offensive efficiency. With 1898 assists per season, their passing is on par with the Celtics. They have 516.62 steals per season, the lowest among the five teams, indicating weaker defensive pressure on ball handlers. The team commits 1093 turnovers per season, which is quite high. Their foul count (1543.25) is the highest among the teams, suggesting aggressive defensive play. Brooklyn has a high number of disqualifications (9.375 per season) and 379.87 blocks, suggesting they rely heavily on shot-blocking defense. Brooklyn Nets have moderate shooting efficiency and strong rebounding, but they lack in defensive pressure (steals) and they also need to reduce fouls and turnovers. Charlotte has the lowest FG% (43.7%) and FT% (76.33%) in this study, indicating inefficiency in shooting and scoring opportunities. With 1849.5 assists per season, they have the lowest passing efficiency among the teams. They record 556.8 steals per season, which is moderate. Their turnover count (1029.12 per season) is moderate, and their personal fouls (1406.5) are the lowest, suggesting a disciplined defense. Charlotte has 8.25 disqualifications per season and 353.75 blocks, the lowest blocking count in the dataset. The Hornets struggle with scoring efficiency, passing, and blocking but commit fewer fouls than other teams. They need to improve offensive execution and defensive presence. The Chicago Bulls have moderate FG% (44.03%) and FT% (78.73%), making them a decent offensive team. Chicago records 1843.87 assists per season, the lowest in this dataset, indicating weak ball movement. The Bulls have 580.12 steals per season, the highest in this study, showing excellent defensive pressure. They commit 1047.75 turnovers per season, which is moderate compared to other teams. The Bulls have 7.75 disqualifications per season and the lowest block count (321.5), suggesting a lack of rim protection. The Chicago Bulls have a strong defensive presence (high steals), but their shot-blocking and playmaking need improvement. Hence from the descriptive analysis of each variable for sample team regular seasons key takeaways and comparison has been depicted in table 2

Table 2 Key Takeaways and Comparison (Regular seasons games)

Team	Strength	Weaknesses
Atlanta Hawk	Strong passing, reliable at free throws and active on defense	Inefficient shooting, frequent fall-outs, prone to losing possession and weak rim protections
Boston Celtic	Most efficient shooting team, Strong free- throw shooting, Most disciplined team, Best rim protection, Strong ball-handling.	Moderate assists and lower steals
Brooklyn Net	Moderate FG% and good ball movement	Poor free throw efficiency, aggressive but undisciplined defence, players frequently fallout and passive defence
Charlotte Hornets	Disciplined defence, decent defensive play	Struggles with shot accuracy, Worst free-throw shooting team, Poor ball distribution, Lacks rim protection.
Chicago Bulls	Most aggressive defence and decent FG%	Poor team coordination and weakest rim protection

5. Conclusion and Suggestions

The study analysed the performance metric of NBA teams for 8 regular seasons. The top 5 team was selected from the official NBA website (*Cume Stats | Stats | NBA.com*) in an alphabetical order and different variables like field goal percentage (FG%), free throw percentage (FT%), assists (AST), personal fouls (PF), disqualifications (DQ), steals (STL), turnovers (TO), and blocks (BLK) was analyzed and compared (Kotzias, “The Four Factors of Basketball as a Measure of Success”). It was found out that while most performance metrics remained consistent across teams, free throw efficiency and shot-blocking ability varied significantly, affecting overall team performance. The team with lower FT% can focus on free throw drills, biomechanical analysis and can enhance mental conditioning of the players. Teams with lower BLK like Chicago Bulls can focus on rim protection training and positional awareness drills for tall men. Turnovers did not show significant differences, but teams like Charlotte Hornet can prioritize ball security through structured offensive play, decision making training and passing accuracy drill

This study is limited to only descriptive analytics of 5 NBA teams for the past 8 years only for the regular season. Future study can include more advanced data analysis techniques for predicting players and team performance based on individual shooting mechanics, fatigue levels, and game conditions. An extension of this study can also include more variables like three-point percentage (3P%), effective field goal percentage (eFG%), and defensive rating (DRtg) that can provide a more in depth understanding of the team performance. This study can be referred to by policy makers, trainers, researchers to have an insight for team management for optimizing roster decisions, developing game strategies to identify focus areas in training.

Works Cited

- Cao, et al. "The Effects of High-Intensity Interval Training On Basketball Players: A Systematic Review and Meta-Analysis." *Journal of Sports Science and Medicine*, vol. 24, 2025, pp. 31–51.
- Cume Stats | Stats | NBA.com. www.nba.com/stats/cumestats?TeamID=1610612737. Accessed 22 Feb. 2025.
- Fei Lu, and Hao Li. "Psychological factors in training of basketball players to improve their shooting accuracy." *Mobile Information Systems*, vol. 2023, Aug. 2023, pp. 1–10.
- França, Cíntia, et al. "A kinematic analysis of the basketball shot performance: impact of distance variation to the basket." *Acta of Bioengineering and Biomechanics*, vol. 24, no. 1, Jan. 2022, <https://doi.org/10.37190/abb-01995-2021-06>.
- "Free Throw Percentage - (Sports Journalism) - Vocab, Definition, Explanations | Fiveable." Fiveable, library.fiveable.me/key-terms/sports-journalism/free-throw-percentage. Accessed 14 Apr. 2025.
- García, Franc, et al. "Differences in Physical Demands between Game Quarters and Playing Positions on Professional Basketball Players during Official Competition." *PubMed*, vol. 19, no. 2, June 2020, pp. 256–63. pubmed.ncbi.nlm.nih.gov/32390718.
- Gutiérrez. "Performance analysis in basketball players." *International Journal of Sports, Exercise and Physical Education*, vol. 6, no. 1, 2024, pp. 86–87.
- Horvat, Tomislav, et al. "A Data-Driven Machine learning algorithm for predicting the outcomes of NBA games." *Symmetry*, vol. 15, no. 4, Mar. 2023, p. 798. <https://doi.org/10.3390/sym15040798>.
- I. Andrianova, et al. "Specific features of 3×3 basketball: factor analysis of the key performance indicators and their impact on game performance in the elite leagues." *Journal of Physical Education and Sport*, vol. 22, no. 10, 2022.
- Kelmendi, Diellza Shaqiri, et al. "Kinematic analysis of the basketball free throw in preparation phase of elite athletes." *International Journal of Human Movement and Sports Sciences*, vol. 9, no. 6, Nov. 2021, pp. 1204–12. <https://doi.org/10.13189/saj.2021.090614>.
- Kotzias, Konstantinos. "The Four Factors of Basketball as a Measure of Success." *Statathlon*, 17 Sept. 2023, statathlon.com/four-factors-basketball-success. Accessed 14 Apr. 2025.
- Kumar. "Physiological variables and their performance of basketball players." *International Journal of Yogic, Human Movement and Sports Sciences*, vol. 4, no. 1, 2019, pp. 1475–76.
- Liang, Zeyu, et al. "Data-driven insights into basketball performance: Unveiling the impact of advanced analytics on player and team efficiency." *Molecular & Cellular Biomechanics*, vol. 22, no. 1, Jan. 2025, p. 1083. <https://doi.org/10.62617/mcb1083>.
- Liu, Xinhao. "Analysis on the Impact of Foreign Players on the Chinese Basketball Association from the Perspective of the Sports Economy." *Advances in Economics Management and Political Sciences*, vol. 24, no. 1, Sept. 2023, pp. 251–55. <https://doi.org/10.54254/2754-1169/24/20230446>.
- M. Libed, et al. "Basketball Game Analysis using Naïve Bayesian Classification Algorithm." National University, Philippines, 2017.

- Özer, Ömer. "Investigation of Strength, Flexibility and Balance Parameters with Performance Dimension in Basketball Players." *Journal of Education and Learning*, vol. 8, no. 5, Sept. 2019, p. 225. <https://doi.org/10.5539/jel.v8n5p225>.
- Öztürk, Yasemin, et al. "Gender Differences in Sports Participation: A Multi-Level Analysis." *Bogazici Journal*, vol. 35, no. 2, Dec. 2021, <https://doi.org/10.21773/boun.35.2.3>.
- Pan, Hang, et al. "Biomechanical analysis of shooting performance for basketball players based on Computer Vision." *Journal of Physics Conference Series*, vol. 2024, no. 1, Sept. 2021, p. 012016. <https://doi.org/10.1088/1742-6596/2024/1/012016>
- Pang, Haolin, et al. "Research on the Main Psychological Factors influencing basketball players' Athletic performance: the importance of psychological Quality." *International Journal of Early Childhood Special Education (INT-JECSE)*, vol. 29, no. 5, Jan. 2020, p. 491. <https://doi.org/10.24205/03276716.2020.1046>.
- Pratama, Rryan. "THE EFFECT OF BLOCK PRACTICE, SERIAL PRACTICE AND RANDOM PRACTICE TO IMPROVE BASKETBALL FUNDAMENTAL SKILL FOR BEGINNER." *The 1st Yogyakarta International Seminar on Health, Physical Education, and Sports Science 2017*, Oct. 2017.
- Raptors. "Field Goal Percentage (FG%) - NBA.com: Jr. NBA." *NBA.com: Jr. NBA*, 13 Oct. 2015, jr.nba.com/field-goal-percentage-fg.
- Richwin Dmello. "THE IMPACT OF DATA ANALYTICS ON PLAYER PERFORMANCE IN PROFESSIONAL SPORTS A SYSTEMATIC REVIEW." *International Research Journal of Modernization in Engineering Technology and Science*, vol. 5, no. 4, 2023.
- Sarlis, and Tjortjis. "Sports Analytics — Evaluation of basketball players and team performance." *Information Systems*, vol. 93, 2020.
- Sarlis, Vangelis, and Christos Tjortjis. "Sports Analytics: Data mining to uncover NBA player position, age, and injury impact on performance and economics." *Information*, vol. 15, no. 4, Apr. 2024, p. 242. <https://doi.org/10.3390/info15040242>.
- Shamim, Parvez. "The impact of sitting height on basketball performance: An analysis of scoring, rebounding, blocking, and assisting abilities." *RESEARCH REVIEW International Journal of Multidisciplinary*, vol. 5, no. 1, Jan. 2020, pp. 241–46. <https://doi.org/10.31305/rrijm.2020.v05.i01.050>.
- Shao, Xinbo, and Yuwei Sun. "A study on the impact of basketball on the physical fitness and health of adolescents based on the method of correlation analysis." *Journal of Environmental and Public Health*, vol. 2022, no. 1, Jan. 2022, <https://doi.org/10.1155/2022/6520518>.
- SOFYAN, et al. "In-Depth Analysis of Exercise and Impact to Basketball Athletes Performance from A Bibliometric Perspective." *Mediterranean Journal of Sport Science*, vol. 5, no. 4, 2022, pp. 785–98.
- Struzik, Artur, et al. "Biomechanical analysis of the jump shot in basketball." *Journal of Human Kinetics*, vol. 42, no. 1, Oct. 2014, pp. 73–79. <https://doi.org/10.2478/hukin-2014-0062>.
- Supola, Bence, et al. "The Role of Secondary Assists in Basketball – an Analysis of Its Characteristics and Effect on Scoring." *International Journal of Performance Analysis in Sport*, vol. 22, no. 2, Feb. 2022, pp. 261–76. <https://doi.org/10.1080/24748668.2022.2039090>.

- Tsagris, et al. "On predicting an NBA game outcome from half-time statistics." *Discover Artificial Intelligence*, vol. 2024, 2024, pp. 1–23. Tsiannis, Angelos, and Christodoulos Efstathiades. "Towards Predicting Basketball Player Positions with Transformative Insights." *Machine Learning and Knowledge Extraction*, vol. 7, no. 1, Feb. 2025, p. 11. <https://doi.org/10.3390/make7010011>.
- Tyshchenko, Valeria, et al. "Factor analysis of indicators of physical and functional preparation of basketball players." *Journal of Physical Education and Sport*, Jan. 2018, repository.ldufk.edu.ua/bitstream/34606048/15134/1/%d0%a2%d0%b8%d1%89%d0%b5%d0%bd%d0%ba%d0%be_%d0%93%d0%bd%d0%b0%d1%82%d1%87%d1%83%d0%ba_%d0%9f%d0%b0%d1%81%d1%96%d1%87%d0%bd%d0%b8%d0%ba_%d0%91%d1%83%d0%b1%d0%b5%d0%bb%d0%b0_%d0%a1%d0%b5%d0%bc%d0%b5%d1%80%d1%8f%d0%ba_2018.pdf.
- Yao, Peng. "Real-Time analysis of basketball sports data based on deep learning." *Complexity*, vol. 2021, no. 1, Jan. 2021, <https://doi.org/10.1155/2021/9142697>.
- Zhang. "Impact analysis of basketball exercise strength based on machine learning in the mental health of college students." *Computational Intelligence and Neuroscience*, vol. 2022, 2022, pp. 1–9.
- Zheng, Lei, et al. "Predicting scoring trends in basketball games based on multivariate time series analysis." *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Nov. 2023, <https://doi.org/10.2478/amns.2023.2.01228>.
- Zuccolotto, et al. "Spatial performance analysis in basketball with CART, random forest and extremely randomized trees." *Annals of Operations Research*, vol. 2023, no. 325, 2023, pp. 495–519.

Relationship Between Nighttime Sleep, Daytime Naps, and GPA By Yusuf Abdellatif

Abstract

Sleep plays a critical role in adolescent cognitive development, yet biological shifts often lead to insufficient rest among high school students. This study investigated the relationship between nighttime sleep duration, daytime napping, and academic performance (GPA). Data was collected via an anonymous online survey from 84 high school students (grades 9–12), assessing self-reported sleep habits and GPA. Participants were categorized into three categories: non-nappers, short nappers (<1 hour), and long nappers (>1 hour).

Results indicate that the correlation between sleep and GPA is heavily influenced by napping behavior. Among students who did not nap, results aligned with traditional expectations: those obtaining more than 8 hours of sleep achieved the highest average GPAs. However, in the short-nap group, students sleeping only 5–6 hours at night outperformed those sleeping longer, suggesting a compensatory cognitive benefit from brief naps. Students taking naps longer than one hour generally showed lower academic performance compared to other groups. The findings suggest that while extended nighttime sleep is ideal, short daytime naps may serve as an effective strategy to maintain academic performance during periods of reduced sleep, whereas excessive napping is associated with poorer educational outcomes.

Introduction

One of the most crucial and fundamental biological processes that plays a major role in brain development, memory, and cognitive function is sleep. Naturally we try to gain the appropriate amount of sleep that our bodies need. However, during adolescence, changes in circadian rhythm cause teenagers to experience a natural delay in sleep, thus making teenagers wake up later and sleep later. This results in insufficient sleep patterns on school nights which demands early wake-ups with early bed times. Previous research has demonstrated that chronic sleep deprivation in adolescents is associated with decreased attention, impaired memory, and reduced academic performance. Given the importance of memory, learning, and critical thinking, all heavily impacted by sleep duration, examining the relationship between sleep and academic performance in high school students is essential for understanding how sleep habits may influence education outcomes and overall adolescent health.

Research Question

This paper seeks to answer the question of how nighttime sleep duration is associated with academic performance in high school students, and what is the extent to which daytime napping modifies the relationship.

Method

This study used a survey design to examine the relationship between nighttime sleep duration and academic performance in high school students. Participants were high school students ranging from grades 9-12 recruited through peer and social networks, and participation was voluntary and anonymous. No identifying information was collected.

Materials and Procedure

Data was collected using an online questionnaire administered through Google Forms. The survey included the following questions asking students to self assess their average sleep duration, daytime nap duration if any, GPA, Gender, if they felt tired during school days, and if they used electronic devices an hour before bedtime. GPA ranges were later converted into numerical midpoint values to allow for quantitative analysis. All metrics were self assessed including GPA.

When examining the effect of daytime nap duration on academic performance, students were placed into 3 different nap groups; students who didn't take naps, students who took naps for less than 1 hour, and students who took naps for more than one hour. GPA was then compared for each of these nap groups.

Results

The survey responses from 84 high school students were analyzed to examine patterns between nighttime sleep duration, daytime napping behavior, and academic performance. The majority of responses were from upperclassmen.

When examining the overall sleep vs GPA result (Figure 1), students reporting 6–7 hours of sleep had the highest average academic performance, while students reporting more than 8 hours of sleep had the lowest average academic performance.

In the no-nap group (Figure 2), no students reporting less than 5 hours of sleep reported taking an after-school nap. Results from this group show that students who slept for more than 8 hours had the highest average GPA, while students who slept for 5–6 hours had the lowest average GPA.

In the less than one-hour nap group (Figure 3), students who slept for 5–6 hours had the highest average GPA, whereas students who slept for more than 8 hours had the lowest average GPA.

In the more than one-hour nap group (Figure 4), no students who slept for more than 8 hours reported taking naps longer than one hour. Within this group, students reporting less than 5 hours of nighttime sleep had the highest average GPA, while students reporting 7–8 hours of sleep had the lowest average GPA. Overall, students who reported naps longer than one hour had lower average GPAs compared to students in the other nap groups

Discussion

The results of this study indicate that the relationship between nighttime sleep duration and academic performance in high school students is not uniform and appears to be influenced by daytime napping behavior. When all students were analyzed together, average GPA varied across sleep-duration categories, suggesting that sleep duration alone does not fully explain differences in academic performance.

When students were filtered by nap duration, clearer patterns emerged. Among students who did not take daytime naps, longer nighttime sleep duration was associated with higher average GPA, with students sleeping more than eight hours showing the highest academic

performance. Not a single person who slept for less than 5 hours did not take a nap after school. These findings align with existing research indicating that sufficient, uninterrupted nighttime sleep supports cognitive functions such as attention, memory consolidation, and learning. In this group, nighttime sleep duration may serve as a better indicator of sleep quality and regularity.

However in the less than one hour nap group GPA, the results contrasted the former. Survey shows that students who slept for 5-6 hours had the highest average GPA while students who slept for more than 8 hours had the lowest average GPA. However, the average GPA across all sleep groups never fell below a 4.0 GPA. It may seem the reason for students to possibly have a lower GPA even though they had the full recommended 8 hours and a small nap could be due to them simply not doing their work or being too lazy so they decide to sleep the day away. But that doesn't align with the fact that the more than 8 hour sleep group still had an average of a 4.0 GPA. It is likely that the students who need this nap after taking the recommended sleep needed for their age is due to them having sports or extracurricular activities that they need to attend to, or they are growing up physically so they need extra sleep.

One possible explanation for the relatively high average GPA observed among students sleeping 5–6 hours in the less than one-hour nap group is the compensatory effect of short daytime naps. Brief naps lasting less than one hour have been shown to improve alertness and attention without significantly disrupting nighttime sleep. Students in this group may be using short naps strategically to maintain cognitive performance despite shorter nighttime sleep, particularly if they are highly academically engaged. As a result, short naps may help offset some of the negative effects of reduced sleep duration and allow these students to sustain higher academic performance in the short term.

For the more than one hour nap group the highest GPA was the less than 5 hour group and lowest was the more than 7-8 hour group. No student said that they slept for more than 8 hours and took a more than an hour long nap after school. Sleeping less than 5 hours reduces focus, memory consolidation, and slower cognitive processing. The best way to cope with those losses is to take after school naps. In the less than one hour nap group no data was collected from students who took less than a 5 hour sleep at night. This is because not only did they need to regain the lost sleep, they also were very tired causing them to sleep for more than an hour. There is a possibility that some of these students wake up later in the night such as at 7 or 8 pm. Causing them to do all their work into the midnight and later, which then causes them to miss on sleep, which then results in a negative feedback loop. The average GPA of this nap group was less than the other nap groups. The less than 5 hours of sleep group had the highest GPA because they were the only ones that truly needed it. In all of the other groups, taking a nap for more than an hour is either unnecessary or over napping, especially in the 7-8 hours of sleep group.

Conclusion

This study examined the relationship between nighttime sleep duration and academic performance in high school students, with a focus on how daytime napping affects this relationship. Overall, the results showed that sleep and GPA were related in different ways depending on students' napping habits, suggesting that sleep duration alone does not fully

explain academic performance.

When students were grouped by nap duration, different patterns were observed. Students who did not take naps generally showed higher academic performance with longer nighttime sleep. In contrast, students who took naps showed different patterns. In the less than one-hour nap group, students sleeping 5–6 hours had relatively high GPAs, which may be explained by the benefits of short naps that help improve alertness. Students who took naps longer than one hour tended to have lower average GPAs overall.

These findings suggest that both nighttime sleep and daytime napping should be considered when examining academic performance in adolescents. While some students may use short naps to help manage shorter sleep durations, longer naps may be linked to lower academic performance. Due to the use of self-reported data and the observational design of this study, causal conclusions cannot be made. However, this study highlights the importance of healthy and balanced sleep habits for supporting academic success in high school students.

Figure 1

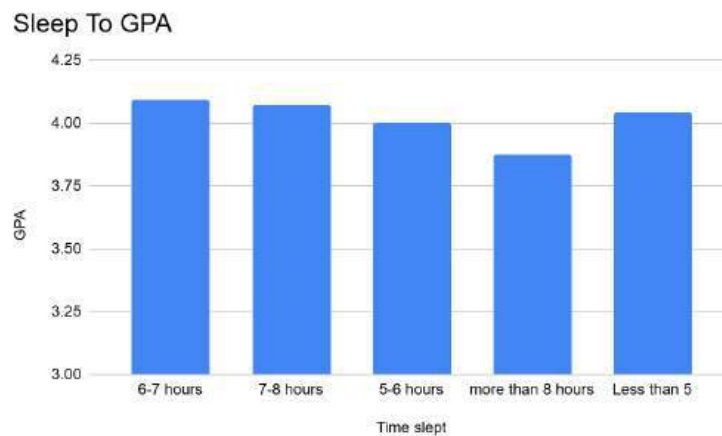


Figure 2



Figure 3

Less Than One Hour Nap Sleep to GPA

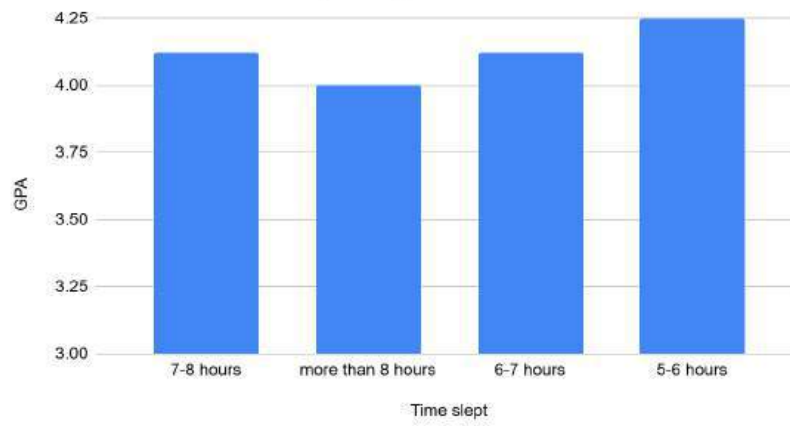
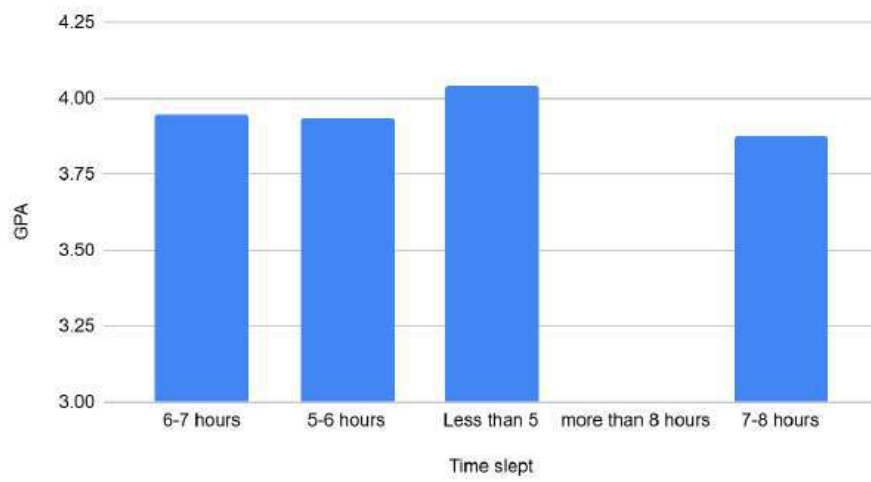


Figure 4

More Than One Hour Nap Sleep to GPA



Work Cited

- Wang, Maohua, et al. "Sleep Duration and Subject-Specific Academic Performance among Adolescents in China." *Npj Science of Learning*, vol. 10, no. 1, 30 Sept. 2025, www.nature.com/articles/s41539-025-00361-y#citeas, <https://doi.org/10.1038/s41539-025-00361-y>.
- Tarokh, Leila, et al. "Sleep in Adolescence: Physiology, Cognition and Mental Health." *Neuroscience & Biobehavioral Reviews*, vol. 70, no. 1, Nov. 2016, pp. 182–188, www.ncbi.nlm.nih.gov/pmc/articles/PMC5074885/, <https://doi.org/10.1016/j.neubiorev.2016.08.008>.
- Lovato, Nicole, and Leon Lack. "The Effects of Napping on Cognitive Functioning." *Progress in Brain Research*, vol. 185, 2010, pp. 155–66, www.ncbi.nlm.nih.gov/pubmed/21075238/, <https://doi.org/10.1016/B978-0-444-53702-7.00009-9>.
- Wheaton, Anne G., et al. "Short Sleep Duration among Middle School and High School Students — United States, 2015." *MMWR. Morbidity and Mortality Weekly Report*, vol. 67, no. 3, 26 Jan. 2018, pp. 85–90, www.cdc.gov/mmwr/volumes/67/wr/mm6703a1.htm, <https://doi.org/10.15585/mmwr.mm6703a1>.

A Quantitative Model on Physical Activity and Learning Readiness among Preschool Children By Muhun Jung

Abstract

This study examines a possible relationship between the frequency of physical activity and learning readiness among preschool-aged children through a simple optimization-based quantitative model. Since early childhood is an important developmental phase for children to develop cognitive and behavioral capacities that support later academic success, it is critical to identify modifiable influences on learning readiness among preschoolers. Using publicly available and nationally representative data from the 2024 National Survey of Children's Health, the analysis in this study focused on children in the range of age from three to five, while modeling learning readiness as a linear function of the frequency of physical activity. The estimated relationship is then interpreted within a context of optimization to identify specific levels of physical activities that may maximize predicted learning readiness under real-world constraints. According to descriptive patterns, there exists substantial variation in both physical activity and learning readiness across the preschool population. Results in this study show a positive association between the frequency of physical activity and the levels of predicted learning readiness. This supports the hypothesis in this study that physical activity contributes to developing early learning capacity. While causal inference is limited by the cross-sectional and parent-reported nature of the data used in this study, the findings in this study provide important quantitative evidence that physical activity on a regular basis may play a complementary role for contributing to developmental preparedness for school. This study ultimately contributes to the literature by showing how a simple quantitative framework may inform early childhood activity planning and decision-making.

Introduction

Early childhood is known to be a critical developmental stage where individuals experience foundational cognitive, behavioral, and neurological capacities to be rapidly formed. During the pre-school years, children tend to have particularly high brain plasticity, making them especially responsive to environmental and behavioral inputs that shape later learning outcomes (Shonkoff and Phillips, 2000). Therefore, it is important to identify modifiable behaviors that support early learning capacity as a main concern in developmental science and education research.

Learning readiness is often referred to as school readiness that describes an ability of a child to effectively engage in structured learning environments. This construct normally includes attention regulation, emotional control, persistence, and basic cognitive skills (Duncan et al., 2007). Sustained academic success and positive behavioral outcomes over time tend to be demonstrated by children who enter school with higher levels of learning readiness (Hair et al., 2006). Therefore, there has been an increasing scholarly attention on factors that influence learning readiness prior to school entry.

One such factor is physical activity. Physical activity has traditionally been emphasized with its role in physical health but has more recently been linked to cognitive and neurological

development. According to prior research, it has been suggested that movement and active play may enhance executive functioning, attention, and self-regulation in young children (Diamond, 2015). Neurobiological studies indicate that increased cerebral blood flow, synaptic plasticity, and the release of neurotrophic factors that support learning and memory are promoted by physical activity (Ratey, 2008).

Empirical studies dealing with preschool-aged children also reported how there existed positive associations between physical activity and behavioral outcomes relevant to learning. Reduced inattention and improved classroom are the examples (Fredewa and Ahn, 2011). Furthermore, population-level research also suggested that children engaging in regular physical activity exhibit higher levels of cognitive functioning compared to less active cohorts (Hillman, Erickson, and Kramer, 2008). All these findings suggest how physical activity may play an indirect but meaningful role in building early learning capacity.

However, most of the existing literature worked on correlational or linear analytical frameworks without considering practical constraints on children's daily routines. Preschool schedules tend to be often limited by time availability, childcare structures, and family circumstances. All of them tend to restrict the amount of physical activity children may realistically engage in. Furthermore, a strictly linear relationship between physical activity and learning outcomes may not fully account for the possibility of diminishing or non-linear effects, such as excessive or insufficient physical activity may reduce the possible benefits (Janssen and LeBlanc, 2010).

In spite of growing interest in the relationship between physical activity and learning, there are only a few studies that analyzed this relationship in the perspective of optimization. Optimization models have been used in the areas of public health and education to identify ideal allocations of limited resources, such as time or intervention intensity (Becker, 1965). Therefore, applying such an approach to preschool physical activity may allow conceptualizing learning readiness as an outcome to be maximized under the real-world constraints.

Seen in this perspective, this study seeks to address this literature gap by applying nationally representative data from the National Survey of Children's Health with a simple framework designed for optimization. This study aims to provide a structured, quantitative approach to identifying an appropriate level of activity in supporting early learning capacity by modeling learning readiness as a function of physical activity frequency among preschool-aged children. To achieve this goal, this study intends to answer the research question about how the frequency of physical activity among preschool-aged children correlates with learning readiness, and what level of activity would maximize learning capacity under real-world constraints. This study hypothesized that preschool-aged children who more frequently engage in physical activity will show higher levels of learning readiness with an optimal range of activity to maximize learning capacity instead of a strictly linear relationship.

Methods

Data Source and Study Sample

In this study, publicly available data from the 2024 National Survey of Children’s Health (NSCH), a nationally representative survey conducted in the United States, was used. In NSCH, parent-reported information on children’s health behaviors and developmental outcomes are collected. To achieve the goal of this study, the analytic sample was restricted to preschool-aged children in an age between 3 and 5, using the child age variable in the dataset. In order to ensure consistency in the analysis, observations with missing responses on main variables were excluded.

Variables

The primary independent variable was physical activity frequency. This variable was measured using a parent-reported item showing how often the child engaged in physical activity during a typical week. This variable was recorded into a numeric scale that higher values represent more frequent activity.

The dependent variable was learning readiness. This variable was used as a proxy for learning capacity. A parent-reported assessment of whether the child was developmentally ready to begin school was used to measure learning readiness. With this measure, cognitive, behavioral, and adaptive characteristics relevant to early learning were captured.

Quantitative Model

To assess the relationship between physical activity and learning readiness, a simple linear model was generated as follows.

$$L = \beta_0 + \beta_1 A$$

Where L was the learning readiness, A was the physical activity frequency, β_0 was the baseline learning readiness, and β_1 was the effect of physical activity on learning readiness. This linear mathematical function was interpreted as an optimization that the objective was to capture the level of physical activity in maximizing the predicted learning readiness as follows.

$$\text{Max}_A^L$$

Where it was subjective to feasible bounds on A showing realistic physical activity levels for preschool-aged children.

Analytical Procedure

To summarize the preschool sample, descriptive statistics were calculated. The linear model was approximated to assess the correlation between physical activity frequency and learning readiness. When interpreting the optimal level of physical activity in maximizing the learning readiness under real-world constraints, the resulting estimates were used.

Results

Sample Characteristics

The final analytic sample consisted of a nationally representative group of preschoolers from the 2024 National Survey of Children’s Health after restricting the dataset to preschool-aged children in an age between 3 and 5 and also removing observations with missing values on main variables. The final analytic sample included children across all main demographic groups and geographic regions in the United States. This supported the generalization of the findings in this study to the broader preschool population.

Descriptive Patterns

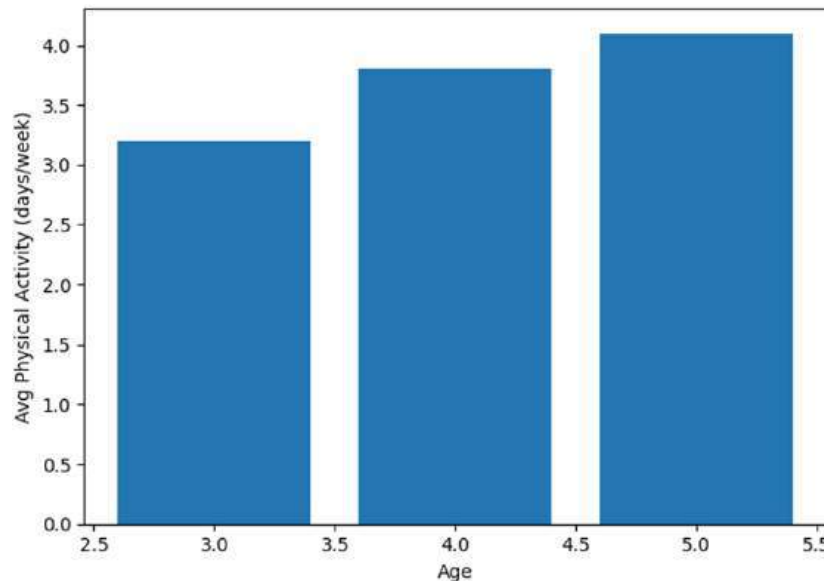


Figure 1. Average weekly physical activity frequency among preschool-aged children (ages 3 to 5)

According to the descriptive analysis shown in Figure 1, a meaningful variation in both physical activity frequency and learning readiness was observed among preschool-aged children. Across the sample, physical activity frequency was recorded in a range from low to high levels, showing how children substantially differed in how often they engaged in active physical activity or exercise during a typical week. Similarly, the scores of learning readiness varied, suggesting how developmental preparedness was different for formal schooling.

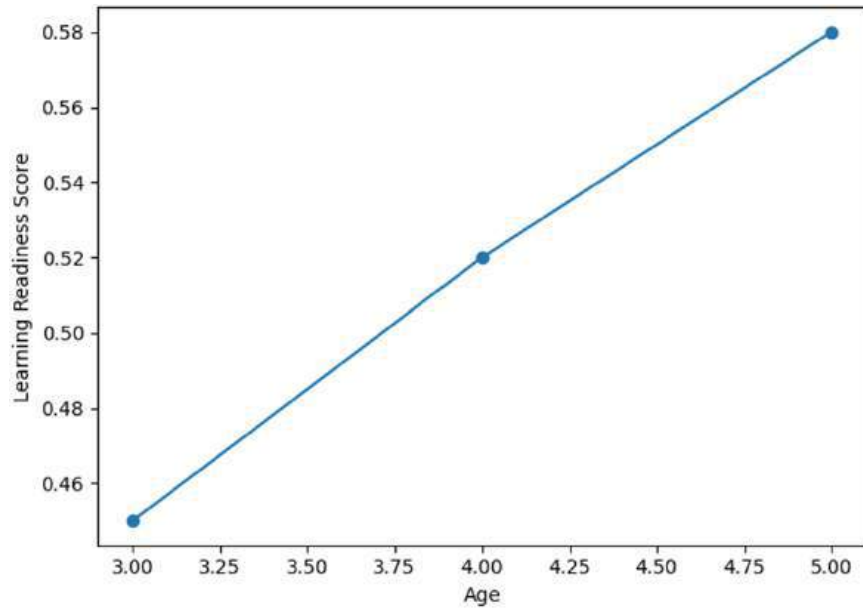


Figure 2. Variation in learning readiness across preschool ages.

When jointly examined, children with higher reported levels of physical activity exhibited higher levels of learning readiness. Preschoolers engaged more frequently in physical activity were more likely to be reported by parents to be developmentally ready to start school when compared to those with lower frequency of physical activity. As shown in Figure 2, this pattern was shown to be consistent across the preschool age range, suggesting that it was not a single age group that drove the aforementioned association.

Regression and Mathematical Model Results

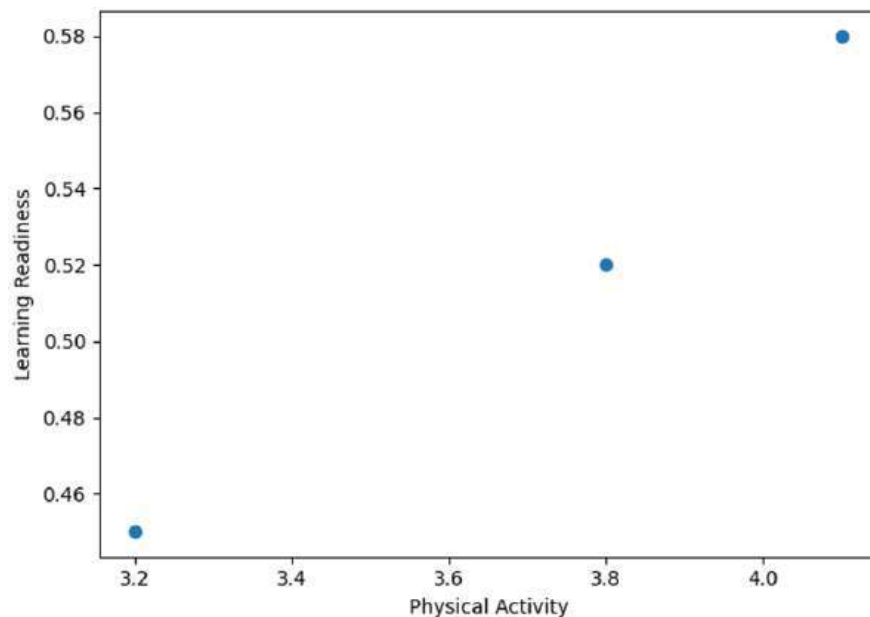


Figure 3. Relationship between physical activity frequency and predicted learning readiness

As shown in Figure 3, learning readiness was estimated by the linear model as a function of physical activity frequency, demonstrating a positive correlation between the two variables. The estimated coefficient of physical activity frequency turned out to be positive, showing how an increase in predicted learning readiness was associated with an increase in physical activity. This finding supports the hypothesis in this study in terms of the positive contribution of physical activity to early learning capacity.

Even if the model was established intentionally simple, it still captured an obvious upward trend in learning readiness as physical activity frequency increased. Interpretable representation of the relationship has been provided by the linear specification, serving as the basis for the optimization interpretation.

Optimization Interpretation

Learning readiness was treated as an objective function to be maximized with the physical activity frequency by using the estimated linear relationship. Applying realistic bounds showing possible activity levels for preschool-aged children, predicted learning readiness turned out to increase as the frequency of physical activity increased within the range of observation. This implies that the more frequently children are engaged with physical activity, the more they improve learning readiness up to practical limits proposed by daily schedules and developmental considerations.

According to the optimization framework, it has been highlighted that physical activity served as a controllable input within early childhood routines. Instead of identifying the association, the model provides a structured way to show how changes in activity frequency lead to changes in learning capacity. While the linear model did not exhibit unlimited benefits, it shows how higher level of physical activity under realistic constraints is associated with higher levels of predicted learning readiness.

Summary of Findings

The results in this study demonstrate how there was a positive relationship between physical activity frequency and learning readiness among preschool-aged children. The findings in this study supported the hypothesis established earlier, showing how physical activity contributes to early learning capacity, and also that an optimal level of physical activity may be identified under real-world constraints. Therefore, these results quantitatively support applying physical activity on a regular basis into preschool routines as a means of developing early learning readiness.

Discussion

This study examined the relationship between the frequency of physical activity and learning readiness among preschoolers by using a simple optimization-based quantitative framework. The results show a positive correlation between physical activity and learning readiness, and this suggests that the more frequently children engage in physical activity, the more they experience higher levels of developmental preparedness for school. These findings turned out to be consistent with the results of prior research that emphasized how movement and

active play serves a role of supporting attention, self-regulation, and early cognitive development.

This study extends existing literature beyond descriptive analyses by interpreting the estimated relationship through a perspective of optimization. The model proposed in this study frames physical activity as a controllable input within daily routines that may be structured to maximize learning capacity in real-world constraints. This approach is specifically relevant for preschool students who may experience limited time and resources that need to be efficiently allocated.

The findings in this study have practical insights especially for parents, early childhood educators, and policymakers. Applying regular, age-appropriate physical activity to schedules of preschoolers may support learning readiness without requiring students to take additional academic instruction. Seen in this perspective, active play may work as a complementary strategy to traditional early learning interventions, while promoting both physical health and cognitive development.

However, this study has several limitations. First, this study relied on parent-reported items that may be subject to reporting bias or measurement errors. Second, the data obtained for this study had cross-sectional nature that may have limited causal interpretation since the observed association did not establish a direct causal relationship between physical activity and improvements in learning readiness. Third, this study generated a simple linear model that did not capture potential non-linear effects or interactions with other factors, including socioeconomic conditions.

Conclusion

Despite aforementioned limitations, this study provides an accessible quantitative framework for examining physical activity and learning readiness of preschool students. The results in this study indicate that students with higher levels of physical activity are more likely to experience greater learning readiness. This suggests that physical activity on a regular basis may support in developing early childhood learning capacity. The findings in this study highlight the potential value of applying age-appropriate physical activity to early childhood schedules as a complement to traditional educational practices. Causality was not established by this study. However, it provides quantitative evidence for physical activity to be linked to early learning preparedness. It is recommended for future study to extend this approach by applying longitudinal data, additional covariates, or non-linear optimization models to further develop estimates of optimal activity levels.

Works Cited

- Becker, Gary S. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press, 1965.
- Diamond, Adele. “Effects of Physical Exercise on Executive Functions: Going beyond Simply Moving to Moving with Thought.” *Child Development Perspectives*, vol. 9, no. 2, 2015, pp. 103–108.
- Diamond, Adele, and Kathleen Lee. “Interventions Shown to Aid Executive Function Development in Children 4 to 12 Years Old.” *Science*, vol. 333, no. 6045, 2011, pp. 959–964.
- Duncan, Greg J., et al. “School Readiness and Later Achievement.” *Developmental Psychology*, vol. 43, no. 6, 2007, pp. 1428–1446.
- Fedewa, Alicia L., and Soyeon Ahn. “The Effects of Physical Activity and Physical Fitness on Children’s Achievement and Cognitive Outcomes.” *Research Quarterly for Exercise and Sport*, vol. 82, no. 3, 2011, pp. 521–535.
- Hair, E., et al. “Children’s School Readiness in the ECLS-K: Predictions to Academic, Health, and Social Outcomes in First Grade.” *Early Childhood Research Quarterly*, vol. 21, no. 4, 2006, pp. 431–454.
- Hillman, Charles H., Kirk I. Erickson, and Arthur F. Kramer. “Be Smart, Exercise Your Heart: Exercise Effects on Brain and Cognition.” *Nature Reviews Neuroscience*, vol. 9, no. 1, 2008, pp. 58–65.
- Janssen, Ian, and Allana G. LeBlanc. “Systematic Review of the Health Benefits of Physical Activity and Fitness in School-Aged Children and Youth.” *International Journal of Behavioral Nutrition and Physical Activity*, vol. 7, 2010, pp. 40–56.
- Magnuson, Katherine A., and Greg J. Duncan. “The Role of Family Socioeconomic Resources in the Black–White Test Score Gap among Young Children.” *Developmental Psychology*, vol. 42, no. 4, 2006, pp. 787–797.
- Ratey, John J. *Spark: The Revolutionary New Science of Exercise and the Brain*. Little, Brown and Company, 2008.
- Shonkoff, Jack P., and Deborah A. Phillips, editors. *From Neurons to Neighborhoods: The Science of Early Childhood Development*. National Academies Press, 2000.
- Strong, William B., et al. “Evidence Based Physical Activity for School-Age Youth.” *The Journal of Pediatrics*, vol. 146, no. 6, 2005, pp. 732–737.
- World Health Organization. *Physical Activity Guidelines for Children Under 5 Years*. WHO Press, 2019.
- U.S. Census Bureau. *National Survey of Children’s Health, Public-Use Data File Documentation*. U.S. Department of Health and Human Services, 2024.

A Comprehensive Review of VEGF and IL-6 Differences in Long COVID Progression Between Sexes By Carissa Chen

Abstract

Long COVID, also known as post-acute sequelae of SARS-CoV-2, presents symptoms like fatigue, cognitive impairment, and cardiovascular issues, with prescribed treatments often targeting symptoms rather than underlying causes. Additionally, sex differences often influence disease presentation, as seen in conditions like heart attacks and Huntington's disease, with emerging evidence suggesting similar differences in Long COVID. Studies indicate women may experience milder acute COVID-19 but have a higher chance of developing Long COVID, while men face more severe acute COVID-19. Vascular endothelial growth factor (VEGF), linked to angiogenesis and COVID-19 severity, and interleukin-6 (IL-6), which plays a vital role in the body's immune response, may play a role in these sex-based differences. This study investigates these two proteins' roles in Long COVID, aiming to clarify their impacts on sex-specific disease progression.

Keywords Long COVID, VEGF, IL-6, sex differences, protein expression

Introduction

While COVID-19 is no longer a public health emergency, the aftereffects of the pandemic continue to impact society. Chief among these is Long COVID, also known as post-acute sequelae of SARS-CoV-2, a condition that occurs after COVID-19 infection. Common symptoms are fatigue and shortness of breath, though others may include loss of sense of smell or taste, memory issues (also known as brain fog), lightheadedness, headaches, chronic cough, chest pain, heart palpitations, anxiety, and depression. Currently, Long COVID remains poorly understood due to its heterogeneity, and treatment primarily focuses on alleviating symptoms. Medications can be used to treat coughs, headaches, anxiety, and depression, and physical therapy can help relieve patients' fatigue and movement issues. However, these treatments often do not resolve the underlying cause of Long COVID, which puts greater importance on the need to understand Long COVID's pathology.

Often, diseases present or progress differently depending on the patient's sex. For instance, men and women experience different symptoms for heart attacks, which has long led to a higher mortality rate of the condition in women. Men usually have the classical symptoms—chest pain and shortness of breath—while women have symptoms not usually associated with heart attacks, such as nausea and vomiting (Schulte et al.). Additionally, certain diseases can have higher frequencies in either males or females. Notably, less than 1% of breast cancer cases occur in men (Ly et al.), and women are much more susceptible to autoimmune diseases such as multiple sclerosis (Voskuhl). Another interesting case is Huntington's disease, an autosomal dominant neurodegenerative disease characterized by excessive CAG repeats in the HTT gene. Some studies have found women with Huntington's disease to present symptoms later than men, and a study found that the disease progressed faster in females (Zielonka and Stawinska-Witoszynska).

Current research on Long COVID has also suggested that sex differences may play a role in its pathology. Farré et al. observed that in a sample of Catalan adults from the COVICAT survey, which assessed the health impact of the COVID-19 on the population of Catalonia, an autonomous community in Spain, women typically experienced milder acute COVID-19 but were more prone to developing Long COVID (Farré et al.). Meanwhile, men tended to exhibit less symptoms of Long COVID but had more severe acute COVID-19. This trend seems to continue in other regions of the world, with a sample of patients in Beijing, China having twice as many male patients with severe acute COVID-19 as female patients (Kong et al.).

Other research has suggested that certain genes are differentially expressed between the sexes during long COVID (Chakraborty et al.; Hamlin et al.), so the proteins they code for might be differentially expressed as well. Based on a comprehensive literature review, we found that vascular endothelial growth factor (VEGF) and interleukin-6 (IL-6) are likely proteins. In brief, VEGF is a family of proteins that is considered the main driver of angiogenesis and lymphangiogenesis, and its potential role in COVID-19 and long COVID has been reported in recent studies (Kong et al.). IL-6 is a cytokine that regulates that functions in a variety of areas, including the acute-phase response, hematopoiesis, and inflammation (Mihara et al.). Many studies have found a possible correlation between elevated IL-6 levels and the severity of acute COVID-19 (Adamescu), as well as between elevated IL-6 levels and long COVID (Yin et al.). Thus, we aim to review these two key proteins, VEGF and IL-6, and their impact on differences in Long COVID progression between sexes.

VEGF Background

VEGF is a family of proteins that is the main driver of angiogenesis and lymphangiogenesis. It consists of five proteins—VEGFA, VEGFB, VEGFC, VEGFD, and placental growth factor (PLGF). Of these, VEGFA, VEGFB, and PLGF regulate angiogenesis, or blood vessel growth, while VEGFC and VEGFD control lymphangiogenesis, the formation of lymphatic vessels (Cozma et al.). These proteins bind to tyrosine kinase VEGF receptors (VEGFRs) to stimulate their associated processes.

VEGFA is the prototype member of the VEGF family in mammals, and it binds to tyrosine kinase VEGF receptors 1 and 2 (VEGFR1 and VEGFR2) in endothelial cells. The latter mediates most of the protein's activities (Ng et al.), including signaling in blood vascular endothelial cells. During embryogenesis, it is expressed in regions of the mesoderm that have been hypothesized to give rise to angioblasts (Apte et al.). After birth, VEGFR2 is expressed on endothelial cells in blood and lymphatic vessels, as well as various other types of cells (Melincovici et al., 2018). Research has also shown that a lack of VEGFR2 leads to embryonic death due to no organized blood vessels being formed (Apte et al.). VEGFR1's functions are comparatively less well-defined, although there is some evidence to suggest that it acts as a regulator of VEGFA activity by limiting the amount available to bind to VEGFR2 (Ng et al.).

VEGFA, encoded by the VEGFA gene on the short arm of chromosome 6 (6p21.1), is perhaps the most important participant in vasculogenesis and angiogenesis. VEGFA and VEGFRs are expressed in regions of blood vessel growth during embryogenesis. Additionally,

Carmeliet et al. and Ferrara et al. separately demonstrated that when a single VEGFA allele is deactivated, it leads to defective vascular development and premature embryonic death (Carmeliet et al.; Ferrara et al.). After birth, VEGFA expression is upregulated in the endothelium of tissues that are still undergoing angiogenesis; besides angiogenesis, the protein also stimulates vascular permeability, which has been shown to play a major role in certain diseases (Apte et al.).

The Role of VEGFs in Various Diseases

In pathological angiogenesis, VEGFA signaling is often hyperactivated. Research has shown that VEGFA plays a role in tumor growth and metastasis, macular degeneration, diabetic retinopathy, rheumatoid arthritis, myocardial ischemia, preeclampsia, and more diseases with vascular components (Melincovici et al., 2018). When VEGFA is secreted by tumor cells, it stimulates the proliferation and survival of endothelial cells, causing abnormal blood vessels to form. Conversely, if VEGFA is targeted via anti-VEGFA antibodies, tumor cell growth decreases (Apte et al.).

In ocular diseases, VEGFA signaling leads to uncontrolled neovascularization. Normally, oxygen levels regulate vasculogenesis in the retina during development; hyperoxia leads to the destruction of the retinal vasculature from dramatic upregulation of VEGFA signaling, while hypoxia promotes vascular development. The increased VEGFA signaling in the former case can lead to the creation of abnormal leaky vessels, which causes intractable fibrosis, eventually leading to retinal detachment and blindness (Apte et al.).

VEGFA also plays a critical role in lung development, and dysfunctional signaling has been associated with multiple lung diseases. In embryogenesis, as the lungs develop, research has found that VEGFA is highly expressed in airway epithelial cells. Additionally, the VEGFA-VEGFR signaling pathway is crucial for normal alveolarization, post-birth alveolar maturation, and development of the vascular bed (Barratt et al.). In adult lungs, VEGFA continues to be highly expressed (**Figure 1**), especially in the alveolar epithelium, although smooth muscle cells, macrophages, endothelial cells, and fibroblasts still have notable levels of expression.

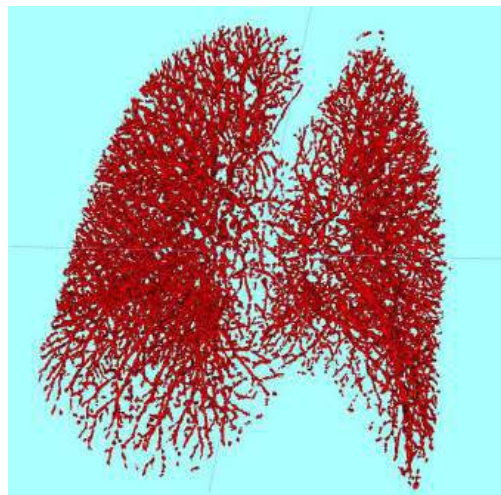


Figure 1. High levels of vasculature in the lung indicate that VEGF plays a key role in respiratory diseases and symptoms. This image is open access and was obtained from the following website:
<https://www.rsipvision.com/lung-vessel-segmentation/>.

Pre-clinical models have found that both inhibition and overexpression of VEGFA in the lungs result in emphysema, a condition that results from alveolar damage and causes shortness of breath (Barratt et al.). VEGF is also associated with other lung diseases, such as acute respiratory distress syndrome (ARDS), which has been shown to be related to COVID-19 pathology. Patients with ARDS had increased VEGF expression in their blood plasma, as well as increased VEGF production in peripheral blood mononuclear cells. It has been hypothesized that increased plasma VEGF levels in ARDS patients might lead to increased vascular permeability in the lungs, directly increasing inflammation and causing the alveolar flooding characteristic of ARDS (Voelkel et al.).

The COVID-19 pandemic produced some new challenges for the lungs. As the virus primarily affects the respiratory system, it is usually associated with endothelial damage and vascular complications, including pulmonary embolism and ARDS. High levels of VEGFA have been found to be associated with poor COVID-19 outcomes (Bahreiny et al.). VEGFA signaling promotes vascular permeability; high levels of it disrupt endothelial cell junctions and increase vessel leakage. In turn, this leads to the development of pulmonary edema and ARDS in COVID-19 patients, causing severe respiratory failure (Bahreiny et al.). Furthermore, since VEGFA promotes immune cell recruitment and stimulates the production of pro-inflammatory cytokines, it directly contributes to the cytokine storm and multi-organ dysfunction found in patients with severe COVID-19. Overall, excessive VEGFA signaling leads to dysfunctional angiogenesis and abnormal vascular development, causing fibrotic changes in the lung and reducing long-term respiratory function in survivors of acute COVID-19 (Bahreiny et al.).

VEGF's Role in Long COVID

During Long COVID, VEGFA plays an important role as well. Multiple papers have reported a connection between VEGFA levels and long COVID progression. For one, VEGFA levels have been reported to be significantly overexpressed in Long COVID patients compared to patients who have never had acute COVID or Long COVID (Farré et al.). Elevated VEGFA levels in the lung epithelial lining may play a role in endothelial barrier dysfunction and increasing vascular permeability in acute respiratory distress syndrome (ARDS) (Kitselman A et al.). Some studies report that Long COVID patients experiencing shortness of breath have higher VEGFA levels as well (Alfadda et al.). Additionally, evidence shows that increased VEGFA signals can lead to worse long COVID symptoms. Neurofilament light protein (NfL), VEGFA, and low hemoglobin levels all contribute to stronger symptoms that cause the patients more discomfort (Lai et al.).

Sex Differences in VEGF Expression During Long COVID

Just like other diseases show differences in symptom and progression between sexes, so does Long COVID. A study on adults in Catalonia regarding Long COVID found that higher

levels of VEGFA were expressed in women than in men, making it one of five proteins that the researchers found were differentially expressed between the two sexes. In fact, VEGFA was the top-ranked protein in the researchers' network analysis and had a statistically significant sex interaction (**Figure 2**). VEGFA had a positive Log2 Fold Change, indicating that the protein is overexpressed in patients with long COVID compared to patients who never had long COVID; furthermore, VEGFA had the most significant p-value of all the proteins that the researchers analyzed (Farré et al.). Thus, VEGFA's role in long COVID is worth further research.

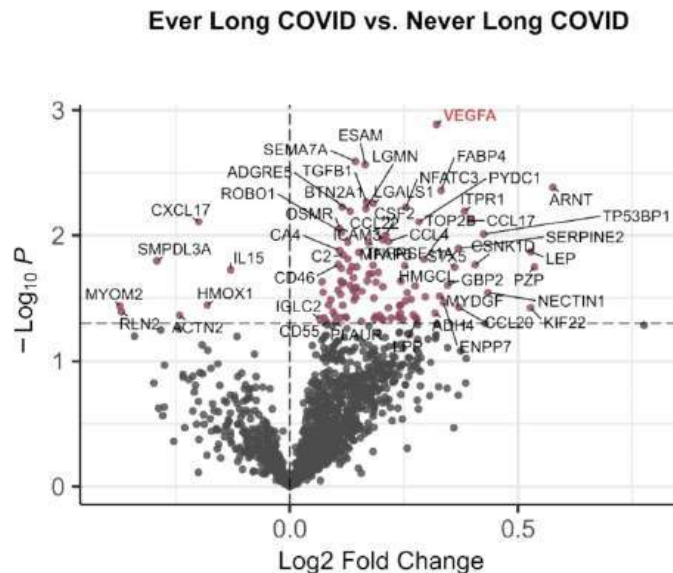


Figure 2. A volcano plot by Farré et al. showing the differential expression of proteins between patients with and without Long COVID. As a notable example, VEGFA is colored in red, and other important proteins are highlighted. This figure has been reproduced under the Creative Commons

Attribution-NonCommercial-NoDerivatives 4.0 International License. The license can be found here: <https://creativecommons.org/licenses/by-nc-nd/4.0/>. No changes were made to this figure.

Besides comparing VEGFA expression for patients with and without Long COVID, the researchers also investigated sex-specific protein interactions. The proteins C2, TSPAN1, M6PR, and DENND2B, and VEGFA all had significant p-values, indicating a significant interaction with sex. However, VEGFA was the only protein whose expression levels did not converge when comparing patients with and without Long COVID. Females with Long COVID had increased VEGFA expression compared to males with Long COVID (**Figure 3**). Additionally, the paper found that menopausal status may influence VEGFA expression in Long COVID, since the analysis they performed found the VEGFA was only differentially expressed among postmenopausal women (Farré et al.).

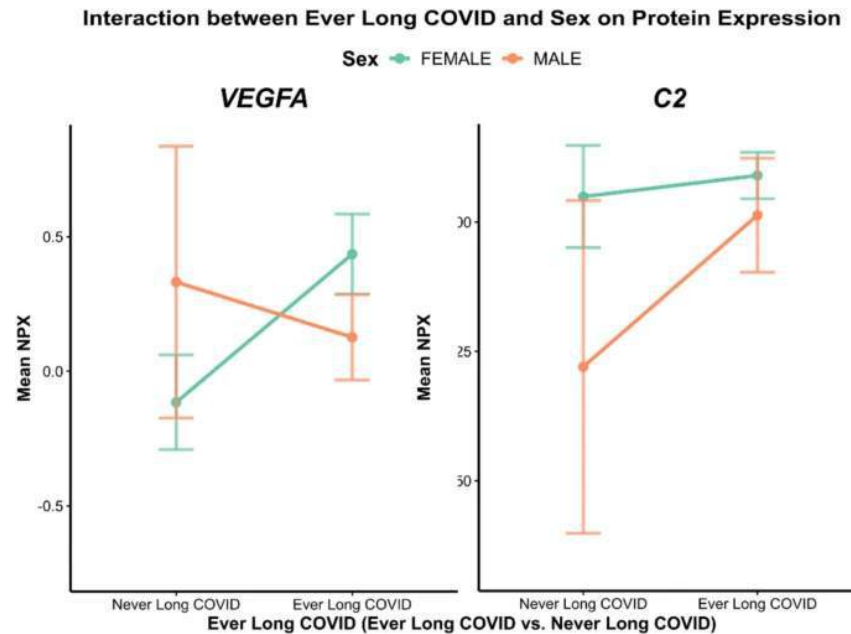


Figure 3. An interaction plot by Farré et al. showing sex differences in C2 and VEGFA expression between patients with and without Long COVID. The two proteins had significant sex-specific differences in expression when comparing the two categories of patients. The Mean Normalized Protein Expressions for VEGFA and C2 differed between sexes for both categories of patients. However, in Long COVID patients, C2 expression was similar between males and females, while relative VEGFA levels were reversed between the two patient categories. This figure has been reproduced under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. The license can be found here: <https://creativecommons.org/licenses/by-nc-nd/4.0/>. No changes were made to this figure.

Females with Long COVID showed higher degrees of endothelial dysfunction than males with Long COVID (Kitselman A et al.). As VEGFR2 is expressed after birth in endothelial cells, and VEGFA primarily binds to VEGFR2, dysfunctional VEGFA signaling may contribute to this observation.

Overall, few studies specifically address sex differences in Long COVID progression and VEGF. VEGFA was the most widely studied VEGF protein in relation to Long COVID, so further research on other members of the VEGF family, such as VEGFB or VEGFD, can be considered.

IL-6 Background

IL-6 is a cytokine that has a variety of functions, especially in the immune system. It promotes the activation of T cells and the differentiation of B cells, and regulates the acute phase response of the immune system (Hunter and Jones, “IL-6 as a Keystone Cytokine in Health and Disease”). As IL-6 is a pleiotropic protein, it also affects various other aspects of the body, such as lipid metabolism, insulin resistance, and the neuroendocrine system.

To carry out activities in the human body, IL-6 binds to a receptor complex that is composed of a receptor subunit, IL-6 receptor (IL-6R), and a signal-transducing subunit, gp130 (also known as CD130). The latter also acts as a receptor for various other proteins and is

essential for normal development, hematopoiesis, and cell survival and growth. Accordingly, it is expressed on both immune cells and non-immune cells, while IL-6R is largely only expressed on hepatocytes, leukocytes, and megakaryocytes (Hunter and Jones, “IL-6 as a Keystone Cytokine in Health and Disease”).

IL-6 signaling is usually performed in two ways: classical IL-6 receptor signaling and trans IL-6 receptor signaling. Classical signaling mediates the protein’s pro-inflammatory activities, while trans signaling controls anti-inflammatory activities (Shekhawat et al.). In classical signaling, IL-6 binds to a membrane-bound IL-6R subunit; as a result, this process only occurs in cells that express IL-6R. The acute phase response, glucose metabolism, hematopoiesis, and the neuroendocrine system are all regulated via classical signaling (Hunter and Jones, “IL-6 as a Keystone Cytokine in Health and Disease”). On the other hand, trans signaling is when soluble IL-6 binds to secreted IL-6R. This forms a complex that increases circulatory IL-6’s half-life, makes it more available throughout the body, and increases the types of cells that are affected by IL-6. Trans signaling is important for the recruitment and apoptosis of leukocytes, the maintenance of effector T cell function, and the inflammatory activation of stromal tissues (Hunter and Jones, “IL-6 as a Keystone Cytokine in Health and Disease”).

The Role of IL-6 in Various Diseases

IL-6 trans signaling is the main cause of IL-6 driven pathology and it plays a central role in the progression of chronic and autoimmune diseases (Jones and Jenkins). In rheumatoid arthritis, a chronic autoimmune disease, the immune system causes inflammation of the tissue that lines the joints of the body. As a result, the joints deform; furthermore, the inflammation can damage other parts of the body, such as the skin, eyes, lungs, heart, and blood vessels (Bullock et al.). Dysregulated IL-6 expression has been found to play a key role in these changes (Yoshida and Tanaka). It helps recruit leukocytes to affected joints in rheumatoid arthritis, causing the local inflammation that eventually leads to joint destruction. Additionally, IL-6 stimulates the proliferation and differentiation of synoviocytes and osteoclasts, respectively, which are associated with the development of osteoporosis and bone destruction. Furthermore, it stimulates excess production of VEGF, causing increased angiogenesis and vascular permeability in synovial tissues, which are characteristics found in patients with rheumatoid arthritis (Yoshida and Tanaka).

Various oncogenic processes are regulated by IL-6 as well—it increases the transcription of molecules vital to cell cycle progression and survival, such as the apoptosis inhibitor survivin (also known as BIRC5). Additionally, IL-6 promotes angiogenesis, tumor invasion and metastasis, and chronic inflammation (Jones and Jenkins).

IL-6 also plays a significant role in acute COVID-19, as it can be involved in vascular leakage and amplification of the cytokine storm—a widespread inflammatory immune response present in severe COVID-19—which are characteristic symptoms of severe forms of the disease (Giannitrapani et al.). Additionally, elevated IL-6 expression is one of the strongest predictors of disease prognosis and mortality in COVID-19 patients (Shekhawat et al.). When the SARS-CoV-2 virus infects cells, they undergo apoptosis or necrosis, which triggers an inflammatory response and the production of pro-inflammatory cytokines, including IL-6.

Eventually, the infection may spread to circulating T regulatory cells, causing them to die and induce a hyperinflammatory state. This condition, which results from oversecretion of cytokines and chemokines, has been coined as a cytokine storm, and research has shown that it can develop into ARDS—another complication from severe COVID-19 (Shekhawat et al.). Since Long COVID is defined as a chronic condition that persists after acute COVID infection, IL-6 may play a key role in the progression of Long COVID.

IL-6's Role in Long COVID

Research has found that Long COVID patients exhibited higher levels of pro-inflammatory cytokines and chemokines, which includes IL-6. Additionally, IL-6 remained upregulated in Long COVID patients throughout, which makes it a potential core biomarker of the condition (Lai et al.). In Lai et al.'s paper on biomarkers in Long COVID, 6 of 20 papers that reported biomarkers between patients with and without Long COVID indicated that Long COVID patients showed higher levels of IL-6.

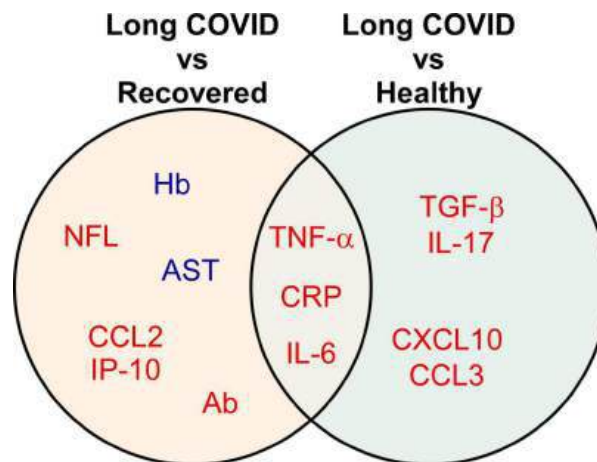


Figure 4. Data from Lai et al.'s review of biomarkers in Long COVID indicates that IL-6 levels are significantly different between people with Long COVID and people who are currently healthy. This figure has been reproduced under the Creative Commons Attribution License (CC BY). The license can be found here: <https://creativecommons.org/licenses/by/4.0/>. No changes have been made to this figure.

Furthermore, evidence has indicated that IL-6 may cause neuron damage in Long COVID, which could mean that IL-6 plays a role in neurological symptoms of Long COVID (difficulty concentrating, “brain fog”, sleep disturbances, etc.) (Lai et al.). Other research has connected IL-6 to yet more symptoms of Long COVID; as cytokines are released during exercise, release of IL-6 may trigger the post-exertional malaise that is often reported by people with long COVID (Low et al.).

In a 2020 study of patients from the COVID Medicine Unit of the University Hospital in Palermo, Italy, higher serum IL-6 levels were associated with twice the risk of having Long COVID, respiratory failure, ARDS, and adverse clinical outcomes (Giannitrapani et al.). However, a study conducted on a Brazilian cohort found that there was no significant difference in IL-6 expression between those with Long COVID and those without (Queiroz et al.). This indicates that more research is needed to ascertain IL-6's role in the progression of Long COVID.

Sex Differences in IL-6 Expression During Long COVID

Besides VEGF, IL-6 has also shown signs of differential expression between sexes in Long COVID. IL-6 is more highly expressed in females with Long COVID than in males (Rudroff). This may correlate with prior findings indicating that females have stronger immune responses and show elevated inflammatory markers (Falahi and Kenarkoohi). Furthermore, IL-6 drives inflammation and is associated with fatigue, poor sleep, depression, pain, and affective disturbances (Rudroff). A Mayo Clinic study found that females made up the highest proportion of Long COVID patients with fatigue-dominant symptoms, and this category of patients expressed the highest IL-6 levels. On the other hand, dyspnea-predominant symptoms were more common in men (Ganesh et al.). Thus, IL-6 may contribute to the higher prevalence of long COVID among females and differing symptom categories between sexes.

As few studies address the relationship between IL-6 and sex differences in Long COVID, more research should be conducted on this topic to gain a more thorough understanding of the disease.

Conclusion

As with many other conditions, Long COVID displays differences in spread and symptom presentation between sexes. During this review, we identified IL-6 and the VEGF family as particularly important proteins in these differences. VEGFA, as the prototype member of the VEGF family, plays a significant role in the lung vasculature of COVID-19 and Long COVID. Female hormones like estrogen stimulate VEGFA expression, explaining the overexpression of VEGFA among female patients. Meanwhile, IL-6 is a key modulator of the immune response, and it drives inflammation in many autoimmune diseases. Females generally have stronger IL-6 reactivity, causing excess inflammation that may lead to the development of Long COVID.

However, there is a limited amount of studies that focus on sex differences in Long COVID. While several papers report that genes such as VEGFA and IL-6 are differentially expressed, understanding of their effects on differences in Long COVID progression remains limited. More research is needed to clarify VEGFA and IL-6's roles in Long COVID's biological mechanism. Additionally, many studies on Long COVID struggle with small sample sizes and geographical limitations, so conducting more comprehensive studies is vital to improve treatment efficacy for both sexes. Overall, investigating the role of proteins in Long COVID development between sexes may lead to breakthroughs that can improve patient health all over the world.

Acknowledgements

I attest that the ideas, graphics, and writing in this paper are entirely my own. I would like to thank Matthew Aronson and Indigo Research for their guidance and expertise.

Works Cited

- Adamescu, Aida-Isabela. "THE ROLE OF IL-6 AND IL-1 IN COVID-19 DISEASE PROGRESSION: THE IMPACT OF GENDER-BASED DIFFERENCES, BMI VARIABILITY AND CO-MEDICATIONS." *FARMACIA*, vol. 73, no. 2, Apr. 2025, pp. 308–16. DOI.org (Crossref), <https://doi.org/10.31925/farmacia.2025.2.6>.
- Alfadda, Assim A., et al. "Clinical and Biochemical Characteristics of People Experiencing Post-Coronavirus Disease 2019-Related Symptoms: A Prospective Follow-up Investigation." *Frontiers in Medicine*, vol. 9, Dec. 2022. Frontiers, <https://doi.org/10.3389/fmed.2022.1067082>.
- Apte, Rajendra S., et al. "VEGF in Signaling and Disease: Beyond Discovery and Development." *Cell*, vol. 176, no. 6, Mar. 2019, pp. 1248–64. www.cell.com, <https://doi.org/10.1016/j.cell.2019.01.021>.
- Bahreiny, Seyed Sobhan, et al. "VEGF-A in COVID-19: A Systematic Review and Meta-Analytical Approach to Its Prognostic Value." *Clinical and Experimental Medicine*, vol. 25, no. 1, Mar. 2025, p. 81. Springer Link, <https://doi.org/10.1007/s10238-025-01583-5>.
- Barratt, Shaney L., et al. "VEGF (Vascular Endothelial Growth Factor) and Fibrotic Lung Disease." *International Journal of Molecular Sciences*, vol. 19, no. 5, Apr. 2018. www.mdpi.com, <https://doi.org/10.3390/ijms19051269>.
- Bullock, Jacqueline, et al. "Rheumatoid Arthritis: A Brief Overview of the Treatment." *Medical Principles and Practice*, vol. 27, no. 6, Sept. 2018, pp. 501–07. Silverchair, <https://doi.org/10.1159/000493390>.
- Carmeliet, Peter, et al. "Abnormal Blood Vessel Development and Lethality in Embryos Lacking a Single VEGF Allele." *Nature*, vol. 380, no. 6573, Apr. 1996, pp. 435–39. www.nature.com, <https://doi.org/10.1038/380435a0>.
- Chakraborty, Chiranjib, et al. "Blueprint of Differentially Expressed Genes Reveals the Dynamic Gene Expression Landscape and the Gender Biases in Long COVID." *Journal of Infection and Public Health*, vol. 17, no. 5, May 2024, pp. 748–66. ScienceDirect, <https://doi.org/10.1016/j.jiph.2024.02.018>.
- Cozma, Angela, et al. "VEGF Polymorphisms (VEGF-936 C/T, VEGF-634 G/C and VEGF-2578 C/A) and Cardiovascular Implications in Long COVID Patients." *International Journal of Molecular Sciences*, vol. 25, no. 16, Jan. 2024, p. 8667. www.mdpi.com, <https://doi.org/10.3390/ijms25168667>.
- Falahi, Shahab, and Azra Kenarkoochi. "Sex and Gender Differences in the Outcome of Patients with COVID-19." *Journal of Medical Virology*, vol. 93, no. 1, 2021, pp. 151–52. Wiley Online Library, <https://doi.org/10.1002/jmv.26243>.
- Farré, Xavier, et al. "VEGFA Sex-Specific Signature Is Associated with Long COVID Symptom Persistence." *BMC Medicine*, vol. 23, no. 1, Oct. 2025, p. 552. Springer Link, <https://doi.org/10.1186/s12916-025-04402-6>.
- Ferrara, Napoleone, et al. "Heterozygous Embryonic Lethality Induced by Targeted Inactivation of the VEGF Gene." *Nature*, vol. 380, no. 6573, Apr. 1996, pp. 439–42. www.nature.com, <https://doi.org/10.1038/380439a0>.

- Ganesh, Ravindra, et al. "The Female-Predominant Persistent Immune Dysregulation of the Post-COVID Syndrome." *Mayo Clinic Proceedings*, vol. 97, no. 3, Mar. 2022, pp. 454–64. www.mayoclinicproceedings.org, <https://doi.org/10.1016/j.mayocp.2021.11.033>.
- Giannitrapani, Lydia, et al. "Can Baseline IL-6 Levels Predict Long COVID in Subjects Hospitalized for SARS-CoV-2 Disease?" *International Journal of Molecular Sciences*, vol. 24, no. 2, Jan. 2023, p. 1731. www.mdpi.com, <https://doi.org/10.3390/ijms24021731>.
- Hamlin, Rebecca E., et al. "Sex Differences and Immune Correlates of Long Covid Development, Symptom Persistence, and Resolution." *Science Translational Medicine*, vol. 16, no. 773, Nov. 2024, p. eadr1032. science.org (Atypon), <https://doi.org/10.1126/scitranslmed.adr1032>.
- Hunter, Christopher A., and Simon A. Jones. "IL-6 as a Keystone Cytokine in Health and Disease." *Nature Immunology*, vol. 16, no. 5, May 2015, pp. 448–57. www.nature.com, <https://doi.org/10.1038/ni.3153>.
- . "IL-6 as a Keystone Cytokine in Health and Disease." *Nature Immunology*, vol. 16, no. 5, May 2015, pp. 448–57. www.nature.com, <https://doi.org/10.1038/ni.3153>.
- Jones, Simon A., and Brendan J. Jenkins. "Recent Insights into Targeting the IL-6 Cytokine Family in Inflammatory Diseases and Cancer." *Nature Reviews Immunology*, vol. 18, no. 12, Dec. 2018, pp. 773–89. www.nature.com, <https://doi.org/10.1038/s41577-018-0066-7>.
- Kitselman A, Kayla, et al. "Sex Differences in Vascular Endothelial Function Related to Acute and Long COVID-19." *Vascular Pharmacology*, vol. 154, Mar. 2024, p. 107250. *ScienceDirect*, <https://doi.org/10.1016/j.vph.2023.107250>.
- Kong, Yaxian, et al. "VEGF-D: A Novel Biomarker for Detection of COVID-19 Progression." *Critical Care*, vol. 24, no. 1, June 2020, p. 373. *Springer Link*, <https://doi.org/10.1186/s13054-020-03079-y>.
- Lai, Yun-Ju, et al. "Biomarkers in Long COVID-19: A Systematic Review." *Frontiers in Medicine*, vol. 10, Jan. 2023. *Frontiers*, <https://doi.org/10.3389/fmed.2023.1085988>.
- Low, Russell N., et al. "A Review of Cytokine-Based Pathophysiology of Long COVID Symptoms." *Frontiers in Medicine*, vol. 10, Mar. 2023. *Frontiers*, <https://doi.org/10.3389/fmed.2023.1011936>.
- Ly, Diana, et al. "An International Comparison of Male and Female Breast Cancer Incidence Rates." *International Journal of Cancer*, vol. 132, no. 8, 2013, pp. 1918–26. *Wiley Online Library*, <https://doi.org/10.1002/ijc.27841>.
- Melincovici, Carmen Stanca, et al. "Vascular Endothelial Growth Factor (VEGF) – Key Factor in Normal and Pathological Angiogenesis." *Romanian Journal of Morphology & Embryology*, vol. 59, no. 2, 2018, pp. 455–67.
- Mihara, Masahiko, et al. "IL-6/IL-6 Receptor System and Its Role in Physiological and Pathological Conditions." *Clinical Science*, vol. 122, no. 4, Oct. 2011, pp. 143–59. *Silverchair*, <https://doi.org/10.1042/CS20110340>.
- Ng, Yin-Shan, et al. "VEGF Function in Vascular Pathogenesis." *Experimental Cell Research*, Special Issue on Angiogenesis, vol. 312, no. 5, Mar. 2006, pp. 527–37. *ScienceDirect*, <https://doi.org/10.1016/j.yexcr.2005.11.008>.

- Queiroz, Maria Alice Freitas, et al. "Cytokine Profiles Associated With Acute COVID-19 and Long COVID-19 Syndrome." *Frontiers in Cellular and Infection Microbiology*, vol. 12, June 2022. *Frontiers*, <https://doi.org/10.3389/fcimb.2022.922422>.
- Rudroff, Thorsten. "Sex-Based Differences." *Long COVID Fatigue: Clinical Sciences, Artificial Intelligence and the Future of Brain Health*, edited by Thorsten Rudroff, Springer Nature Switzerland, 2025, pp. 29–34. Springer Link, https://doi.org/10.1007/978-3-031-89474-9_4.
- Schulte, Kyle J., et al. "Myocardial Infarction Signs and Symptoms: Females vs. Males." *Cureus*, vol. 15, Apr. 2023. www.cureus.com, <https://doi.org/10.7759/cureus.37522>.
- Shekhawat, Jyoti, et al. "Interleukin-6 Perpetrator of the COVID-19 Cytokine Storm." *Indian Journal of Clinical Biochemistry*, vol. 36, no. 4, Oct. 2021, pp. 440–50. Springer Link, <https://doi.org/10.1007/s12291-021-00989-8>.
- Voelkel, Norbert F., et al. "Vascular Endothelial Growth Factor in the Lung." *American Journal of Physiology-Lung Cellular and Molecular Physiology*, vol. 290, no. 2, Feb. 2006, pp. L209–21. journals.physiology.org (Atypon), <https://doi.org/10.1152/ajplung.00185.2005>.
- Voskuhl, Rhonda. "Sex Differences in Autoimmune Diseases." *Biology of Sex Differences*, vol. 2, Jan. 2011, p. 1. PubMed Central, <https://doi.org/10.1186/2042-6410-2-1>.
- Yin, Jing-Xian, et al. "Increased Interleukin-6 Is Associated with Long COVID-19: A Systematic Review and Meta-Analysis." *Infectious Diseases of Poverty*, vol. 12, no. 1, Apr. 2023, p. 43. Springer Link, <https://doi.org/10.1186/s40249-023-01086-z>.
- Yoshida, Yuji, and Toshio Tanaka. "Interleukin 6 and Rheumatoid Arthritis." *BioMed Research International*, vol. 2014, no. 1, 2014, p. 698313. Wiley Online Library, <https://doi.org/10.1155/2014/698313>.
- Zielonka, Daniel, and Barbara Stawinska-Witoszynska. "Gender Differences in Non-Sex Linked Disorders: Insights From Huntington's Disease." *Frontiers in Neurology*, vol. 11, July 2020. *Frontiers*, <https://doi.org/10.3389/fneur.2020.00571>.

Symbols of Conflict: How Wartime Propaganda Misappropriated History and Impacted National Identity in the 20th-Century By Ahriyan Anand

Conflicts often rely on symbols, such as flags or propaganda, to unite the people. Every revolution or governing body has used national symbols to shape identity and public trust, often to serve their own political and ideological interests. From World War II to the revolutions of South Asia and the formation of post-Soviet states, symbols have been exploited as objects of misappropriation and deceit. I argue that revolutionary and wartime movements throughout the 20th century, both historical and contemporary, have consistently employed national symbols not merely to foster unity, but to assert authority and shape public perception in service of their political objectives. This paper proceeds in the following parts: the application of symbols as tools of political power in diverse historical settings throughout the twentieth century. It begins with an investigation of fascist Europe, with particular emphasis upon Nazi Germany and Mussolini's Italy, to demonstrate how fascist regimes exploited historical images, rituals, and myths of the past to instill obedience and justify authoritarian rule. It then proceeds to investigate the application of textiles in Japan, Great Britain, and the United States during wartime, transforming ordinary objects into ideological tools of normalization. It then proceeds with an investigation of South Asia, tracing how religious and cultural symbols, originally used to instill cohesion in anti-colonial movements in India and Pakistan, came to be exploited to justify nationalist and regional conflicts in Bangladesh, Kashmir, and Sri Lanka. It then concludes with an investigation of the post-Soviet era, in which newly independent states in Central Asia, such as Kyrgyzstan, Uzbekistan, Turkmenistan, and Tajikistan, drew on ancient myths and historical figures to establish national identity and compensate for the weakness of their political institutions.

Several scholars have considered the role of ideology, symbolism, and propaganda in moments of conflict and upheaval¹⁰³. Most importantly for this paper, Dunnage, Upadhyay, and Mellon have demonstrated the significance of historical misappropriation in such moments. Mellon's analysis of Central Asia shows how political elites use myth and invented tradition to craft a sense of national belonging that covers up the absence of real democratic authority. While Upadhyay, writing about South Asia, demonstrates something similar but in a different register: instead of myths of origin, it is religion that becomes the primary resource politicians mobilize to secure loyalty and consolidate power. Dunnage's analysis shows that fascist and Nazi regimes did not rely on coercion alone to secure loyalty; they also invested heavily in ritual, symbolism, and tradition-building to cultivate belonging. Finally, Atkins' work on wartime textiles shows how fabric was never just decorative or utilitarian, but a medium through which states inscribed ideology into daily life, turning clothing into a quiet yet constant reminder of duty, sacrifice, and belonging. Upon gaining understanding from the pieces of historical writing, I gained clarity through different types of symbolic misuse and misappropriation across cultures, drawing clearer connections to the examples that led to the reasoning of these pieces.

¹⁰³ Jonathan Dunnage and Nadine Rossol, "Building Ideological Bridges and Inventing Institutional Traditions," *Crime, History & Societies* 19, no. 1 (2015): 90–92.

Nazi Germany

Countries and people use symbols to garner public support and push political and ideological gain during conflict and the reconstruction of any nation.¹⁰⁴ Hitler was no different, using terms like the Supreme race or “Aryan” and the hooked cross as the Third Reich. Perpetuating a common enemy and uniting the disgruntled German people after the punishments by the Allies in the First World War. Furthermore, during the Second World War, Hitler used this same strategy by pushing this supreme race idealistic agenda that spread to German culture and wartime propaganda.¹ It was intertwined with them: Such as the ideals of “Aryan”, the Swastika, and the TotenKopf.¹

The “Aryan” race is a name signaling an extremist or Nazi committing hate crimes against a group of people. However, this was not always the case; the name itself originates in the Indo-Iranian region, as described in Sanskrit, where it refers to a person or group of people of nobility who are well educated. The story goes that these people protected villages in exchange for shelter and were given the name. Hitler didn’t just use the idea of “Aryan” superiority against the Jewish people; he expanded it to target Slavic communities, the Roma, and many others. By labeling these groups as Untermenschen, or subhuman, he created the justification for invading Eastern Europe, forcing millions into slavery, and carrying out mass killings. What might have seemed like a simple conquest was instead framed as a racial war for survival and domination. At the same time, by telling Germans they were the descendants of an ancient and noble “Aryan” race, Hitler instilled in them a sense of pride, destiny, and unity. This myth of being chosen gave the German people a collective identity, tightly bound together under Nazi rule.

The hooked cross, branded across all facets of Nazi life, whether displayed on the flags of the Third Reich or the awards and badges of Hitler Youth and the Military.¹⁰⁵ It was a centerpiece of fascist control and synonymous with their reign of terror. The swastika itself is a part of a greater goal of uniting the people under a common flag. By creating a national symbol, such as a hooked cross. Hitler could create a national identity so strong that it would symbolize loyalty to the established government. The image is deeply rooted in and synonymous with antisemitism, racial hatred, and Nazi ideology.

The SS was a brutal organization designed to implement Nazi Germany’s racial policies and systematic terror into all corners of its control. Fascist regimes usually use such organizations to spread fear and loyalty amongst the citizens. And like most organizations, the SS used national identity symbols to show a sense of nationalism that furthered their agendas in the service of the “country.”¹⁰⁶ A common example of this was the TotenKopf, or death's head, which was famously the SS military badge during the war. A Prussian symbol dating back to the 19th century, symbolizing elite warriors and bearing the motto “Loyalty unto death.” A national symbol from the earliest roots of German history that emphasized a sense of greater purpose to

¹⁰⁴ Simonetta Falasca-Zamponi, *The Aesthetics of Politics: Symbol, Power and Narrative in Mussolini's Fascist Italy* (London: SAGE Publications, 1992): 3-7.

¹⁰⁵ John Dunnage, “Building Ideological Bridges: The Police and the National Community in Italy, 1919–1943,” 2015. 88-90

¹⁰⁶ John Dunnage, “Building Ideological Bridges: The Police and the National Community in Italy, 1919–1943,” 2015 once-decorated. 94-96

the country. Again, the idea of a pure intention is misconstrued as the omen of death associated with SS and their racial enforcements.

Mussolini's Fascist Italy

Fascism is a far-right ideology that puts into focus a strong national pride, a cult-like following, the targeting of minority groups, and the need for expansion. During the Second World War, Mussolini's Black Shirts came to power, establishing the National Fascist Party. Like Hitler, Mussolini used common symbolism, religion, and national pride to unify the people. Creating rituals and imposing propaganda on daily life.¹⁰⁷

Symbolic power in Mussolini's fascist Italy included the fasces, the Roman salute, and Mussolini himself. Ingrained in daily life, these symbols of ancient history and their use as a central, near-divine figure melded ordinary Italian life with the regime, making it synonymous with the regime. The fasces, a once decorated Roman symbol serving as the meaning of justice and control over life and death, were now twisted into the Italian fascist control over its people. Furthermore, the Roman salute common in both Nazi Germany and Mussolini's Italy was twisted even further with the combination of loyalty to a corrupt and tyrannical establishment and symbolizing a strong connection to those groups. In the modern day, it is still seen as a motion that is incredibly racist.¹⁰⁸

Mussolini employed similar tactics through propaganda. By portraying themselves as strong leaders and saviors. A great example is how Mussolini positioned himself as embodying Italy's future and destiny. Enforcing these ideals through parades, speeches. Utilizing developing technology to broadcast his messaging through schools and common places. This complete assimilation blurred the lines between state and citizen, leading to the complete cult-like following that fascism uses for control.

As previously mentioned, Mussolini used events to garner support and loyalty from the Italian people. Some examples of this include the creation of rituals used as symbolic events to instill nationalism and loyalty to the established Raj. An idea mentioned by Falasca-Zamponi was narrative construction, in which Mussolini constructed the noble Roman past with the fascist present and the imperialist future. Through the use of art, film, and radio, the Blackshirts were able to convince the Italian people that their purpose was for the greater good.¹⁰⁹ The symbolism of idealistic history shows how one group can control a nation's path through the reeducation of its citizens. Uniting them under a common history, which can be abused and misrepresented, furthering political gain for the fascist party.

Furthermore, Fascist Italy and Nazi Germany would establish institutions in close relation with the community to emphasize community integration with authoritarian control. Dunnage mentioned that the police were national pillars of the country, working hard to gain prestige in

¹⁰⁷ Simonetta Falasca-Zamponi, *The Aesthetics of Politics: Symbol, Power and Narrative in Mussolini's Fascist Italy* (London: SAGE Publications, 1992). 21-23

¹⁰⁸ Simonetta Falasca-Zamponi, *The Aesthetics of Politics: Symbol, Power and Narrative in Mussolini's Fascist Italy* (London: SAGE Publications, 1992). 78-82

¹⁰⁹ John Dunnage, "Building Ideological Bridges: The Police and the National Community in Italy, 1919-1943," 2015. 112-118

the national community.¹¹⁰ These police forces, known as “black brigades,” were regarded with higher stature than the military and were respected and listened to by all members of the community. The creation of national pillars like the police was used as a symbol for the people to not only rally behind but also show little resistance. Emphasizing the complete control and lack of disobedience against the established government. This allowed authoritarian leaders such as Mussolini to further control the people.

Propaganda through Textiles

Everything can be used as a symbol. Whether it be a fist in the air or the clothes on your back, symbols have shaped the course of human history, leading to an impact on our society's norms and constructions. During wartime struggles, sacrifice and dedication are key qualities that help emerge victorious in the war effort. You may have a large army, but how well it fights depends on the morale of the people in every nation. Thus, the push for normalization began. With the Second World War, Japan, Great Britain, and the United States of America used the textile industry to promote and normalize the war effort, albeit for different reasons.¹¹¹ These superpowers used this industry as a symbol of continued political and ideological control.

Imperial Japan

Emperor Hirohito, like his fellow fascist and imperialistic dictators, established a us vs. them cultural mentality, reinforcing this idea through the harsh treatment of the Chinese and Korean peoples. Children were taught they were the superior asian race, establishing a continued hate towards the entirety of Southeast Asia, conducive to further control entrenched in a sense of great national pride.

The textile industry was one such method. As Japan entered the world stage, its need for continued centralized control relied upon the normalization of the effort against the Southeast and the United States. A fascist dictatorship requires a strong sense of national identity and loyalty. Not to mention the deeper ideological goal of integrating women into the war effort by creating patriotic clothing dedicated to the regime, thereby exemplifying their loyalty and service to national survival.¹¹² The textiles made national identity synonymous with loyalty to the emperor and military pride.⁴ These everyday wear (kimono) and household items led to the normalization of conflict and strengthened the resolve of Imperial Japan as it continued its mass conquest across Asia. The ideological indoctrination of citizens for the political and ideological gain of the Emperor and his controlling party.

Great Britain

Great Britain, one of the last free strongholds against the Nazi terror sweeping through Europe, implored a similar strategy. But rather than ideological indoctrination, sacrifice and

¹¹⁰ John Dunnage, “Building Ideological Bridges: The Police and the National Community in Italy, 1919–1943,” 2015. 1-3

¹¹¹ Jennifer Atkins, *Wearing Propaganda: Textiles on the Home Front in Japan, Great Britain, and America during World War II* (New Haven: Yale University Press, 2005). 34-39

¹¹² Jennifer Atkins, *Wearing Propaganda: Textiles on the Home Front in Japan, Great Britain, and America during World War II* (New Haven: Yale University Press, 2005). 85-86

continued perseverance were emphasized by the new fashions of the time. They were in for the long haul, barely escaping Dunkirk and defending against the Battle of Britain against the waves of Nazi fighters. As England's supply grew smaller, people's clothing became their own propaganda.¹¹³

As supplies dwindled, the standardization of clothing increased, thereby reducing class distinctions and fostering a sense of unity among the British people.¹¹⁴ Instead of the rich vs. the poor, it was the citizens of the isle against the Nazi terror sweeping Europe. Furthermore, patterns sewn represented allied victory and national resistance, which furthered the citizens' morale and continued war effort. Through this continued textile propaganda, the British people were able to survive under harsh conditions with clothing that stifled individuality. The industry allowed for a revitalization of the war effort and the collective unity of the last stronghold of Europe. In the end, signifying the sacrifice and duty of every civilian to do their part against a tyrannical rule.

United States of America

In December 1941, the U.S.A. joined World War II. Still, reeling from the devastating attack of Pearl Harbor, the country needed unity in the face of the growing fear of the Axis. During the war effort, the country used textile motifs, such as patriotic national symbols, to garner support and foster collective unity.

Symbolism through textile motifs often included propagandistic slogans, memorial motifs such as stars, stripes, and eagles, and even military insignia.¹¹⁵ Obviously, these would be used to recruit youth for the war effort and to support it on European soil. By emphasizing an idea Atkins called "patriotic chic," the military could create a stylish association with the war effort. The United States heavily relied on women in domestic roles, linking consumption and fashion to national survival. This collective association of textiles with everyday life created patriotic symbols still prevalent to this day to motivate people into voting and enlisting in the army.

South Asian Post-Colonial Conflict

On August 14 and 15, 1947, two nations were born. A rivalry so polarizing was the direct cause of three wars and mass events of terror. In the shadow of colonial rule, India and Pakistan were born from fighting towards a goal of independence to dominance in the subcontinent. This event started a rivalry that spanned across the Cold War into the modern world.¹¹⁶

I argue that symbolism played a crucial role in South Asian revolutionary movements that sought to resist colonial domination and build a sense of shared identity. Leaders and activists drew on religious imagery, historical figures, and cultural traditions to inspire unity across diverse communities.¹¹⁷ National flags, songs, and mass demonstrations further

¹¹³ Jennifer Atkins, *Wearing Propaganda: Textiles on the Home Front in Japan, Great Britain, and America during World War II* (New Haven: Yale University Press, 2005). 92-94

¹¹⁴ Jennifer Atkins, *Wearing Propaganda: Textiles on the Home Front in Japan, Great Britain, and America during World War II* (New Haven: Yale University Press, 2005). 142 - 145

¹¹⁵ Jennifer Atkins, *Wearing Propaganda: Textiles on the Home Front in Japan, Great Britain, and America during World War II* (New Haven: Yale University Press, 2005). 1-3

¹¹⁶ Yasmin Khan, *The Great Partition* (New Haven: Yale University Press, 2007). 7-10

¹¹⁷ Partha Chatterjee, *The Nation and Its Fragments* (Princeton: Princeton University Press, 1993). 52-53

transformed abstract ideas of freedom into tangible symbols that ordinary people could rally behind. Even amid divisions of class, language, and religion, these shared symbols helped mobilize the population, fueling both the struggle against colonial power and the ongoing effort to imagine a united nation.¹¹⁸ While national rivalry gave birth to new nations, it was cloaked in the greater struggle for dominance in the new world.

Indo-Pak Independence and Interfaith Unity

Within India, there are three major religious groups: Sikhs, Hindus, and Muslims. Although they did not see eye to eye, the people of the region worked together in many facets, regardless of religion, to gain independence from the British Raj.¹¹⁹ While some used peaceful protest, others employed radical tactics to garner support and foster anti-colonial rhetoric. These figures were part of diverse political groups, each with different ideologies but the same ultimate goal: using religious iconography and unity to inspire the Indian people. Through these efforts, they eventually gained independence.

Historical figures

Among the more famous activists were Gandhi and Nehru, founding members of the Indian National Congress who advocated peaceful protest against the British oppressive occupation of India. Vigilante local legends like Baghat and Udham Singh, members of the H.S.R.A. (Hindustan Socialist Republican Army). With many revolutionary leaders becoming martyrs, like Chandra Sekhar Azad, and political activists seeking representation, such as Taj Singh, founder of the Ghadar party.¹²⁰

I argue that these figures and their tactics not only spread anti-colonial rhetoric but also acted as proponents of independence and religious unity. Peaceful protest, an act coined by Mahatma Gandhi in opposition to the British, spread to all corners of the subcontinent. Gandhi, in affiliation with the Indian National Congress (INC), used his status as a Brahmin (the highest Caste in Indian culture) and his massive Hindu following to organize protests against the British Raj.

The use of hunger strikes and truth force mobilized the Indian people to action, willingly dying for the cause, as seen in events such as Jalianwala Bagh. where thousands of Sikh, Hindu, and Muslim protestors of all ages were shot down in cold blood. While Gandhi did not condone the violence the British employed, he stayed firm in engaging in nonviolent retaliation. His dedication to peace and unwavering determination made him a symbol of independence that Indians, regardless of caste or creed, would unite behind.

Another group in the movement of Interfaith Unity in India was the H.S.R.A., a radical organization that relied on bombings, robbery, and vengeance to spread the word of independence and closure to the atrocities of the British Raj. Iconic members included Baghat and Udham Singh. Each played a unique role in the weakening of British colonial rule. Bhagat

¹¹⁸ Anuradha Upadhyay, *Secular Democracies, Governance and Politics of Religion in South Asia* (New Delhi: Routledge, 2018). 124-126

¹¹⁹ Sugata Bose and Ayesha Jalal, *Modern South Asia* (London: Routledge, 2017). 58-63

¹²⁰ Ramachandra Guha, *India After Gandhi* (New York: HarperCollins, 2007). 60-62

Singh struck at the colonial system through train robberies and bombings, aiming to weaken the British financially and expose the vulnerability of their power. Yet his vision went beyond violence—he wrote extensively about equality, secularism, and the need for solidarity across religions, inspiring both Hindus and Muslims to see themselves as partners in the freedom struggle. His martyrdom at a young age turned him into a unifying symbol, especially for students and workers who began to see independence as both achievable and worth sacrificing for.

Udham Singh, on the other hand, built and relied upon an underground network of Indians in England to carry out direct retribution. Most famously, assassinating Michael O'Dwyer, the former lieutenant governor of Punjab, who had defended the Jallianwala Bagh massacre. His act was not seen as an isolated revenge killing, but as a strike for national dignity, reclaiming honor for the thousands massacred in Amritsar. For Indians at home and abroad, his defiance proved that the struggle for freedom had no borders. The Indian cause was global, and that unity across communities and regions was essential.

Chandra Shekhar Azad brought yet another dimension to the fight. Though politically unaffiliated, he led with a fierce commitment to freedom, refusing ever to be captured alive by the British. Through raids, robberies, and armed clashes, he sustained revolutionary momentum during difficult times. Azad's readiness to fight alongside comrades of every background reinforced the idea that independence was a shared national cause, not a movement restricted to any one faith or region. His death in a shootout, true to his vow of remaining "Azad" (free), left behind a legacy of courage that inspired ordinary people to believe in collective resistance.

Collectively, these individuals, with their ideals rooted in non-violence, armed struggle, or ideological shifts, saw their endeavors transcend them. The moral leadership provided by Gandhi, the martyrdom of Bhagat Singh, Udham Singh's quest for justice on a global platform, and, finally, Azad's resolute courage offered something distinct for every citizen to work towards, united yet apart, to leave British colonialism behind and achieve independence for a united India. Their histories weaved together to form this national nostalgic mythology, where every Hindu, Muslim, Sikh, and, in fact, every caste within every citizen, could find themselves a hero.

Parties and Symbols

Political organizations in colonial India often turned to symbols and religious imagery to build solidarity across the country's deeply divided communities. These symbols were not merely decorative; they became shorthand for shared aspirations, linking spiritual traditions with the political demand for freedom. The Indian National Congress (INC), the Hindustan Socialist Republican Association (HSRA), and the Ghadar Party all relied on this symbolic language to cultivate a sense of belonging that transcended regional and religious boundaries.

The INC, which stood at the forefront of the nationalist movement, was particularly adept at weaving symbolism into everyday life. The spinning wheel, or *charkha*, was perhaps its most enduring emblem.¹²¹ On one level, it represented economic self-reliance, calling Indians to

¹²¹ Upadhyay, *Secular Democracies*. 138-140

boycott British goods and revive the indigenous industry. On another, it became a unifying device, a reminder that despite India's vast diversity, ordinary citizens could participate in the same simple act of spinning. Alongside it, the open palm stood as a gesture of peace and solidarity, signaling nonviolent strength in the face of oppression. These symbols, when carried in protests or printed on banners, helped to establish a recognizable national identity, something that millions of Indians, regardless of religion or caste, could rally behind.

The HSRA, though more radical in its methods, also understood the importance of unifying language. Its famous phrase, "Ram, Mohammed, Singh, Azad," captured the essence of interfaith cooperation, naming Hindu, Muslim, and Sikh figures alongside the idea of eternal freedom.¹²² This slogan was not only a call to arms but also a reminder that the struggle for independence demanded the participation of all communities. It resonated so strongly that Udham Singh himself is said to have adopted it before his execution, a final testament to the belief that India's liberation was rooted in unity rather than division.

The Ghadar Party, operating largely from abroad, similarly employed religious and cultural motifs to rally Indians scattered across the globe.¹²³ By invoking imagery familiar to both Hindu and Sikh traditions, the party turned faith into a political language that inspired diasporic Indians to see themselves as part of a collective cause. This blending of sacred and political symbols allowed the movement to bridge distances and foster solidarity among workers, students, and intellectuals living far from India.

Taken together, these uses of symbolism transformed abstract political goals into shared cultural expressions. They allowed people to imagine themselves as part of a single nation long before independence was achieved. More importantly, they created a moral framework for resistance, framing freedom not simply as a political demand but as a duty sanctioned by faith, culture, and history.

The Struggle for Dominance

I argue that due to the post-colonial conflict between India and Pakistan, the surrounding nations would have never been born. As both nations used symbols to advance national rhetoric, helping or hindering the cause of independence for states like Sri Lanka, Jammu and Kashmir, and Bangladesh. As the war for control continued, India and Pakistan waged three mass conflicts that combined to create the independent nation of Bangladesh and the states of POK and India-controlled Jammu Kashmir. These struggles are deeply rooted in religious toleration and a history of mistrust.

Bangladesh

The war over Bangladesh in 1971 had symbols that each side used to justify its actions and assert control.¹²⁴ For Pakistan, the most powerful symbol was the idea of Islamic unity. Leaders in West Pakistan portrayed the country as the guardian of Muslim solidarity, with the crescent and star standing for loyalty to one Pakistan. Any attempt by Bengalis to demand autonomy or recognition of their language was cast as betrayal, symbolically framed as tearing

¹²² Chatterjee, *The Nation and Its Fragments*. 176-179

¹²³ Bose and Jalal, *Modern South Asia*. 203 -205

¹²⁴ Srinath Raghavan, *1971: A Global History* (Cambridge: Harvard University Press, 2013). 213-216

apart the sacred fabric of the nation. Urdu, too, was turned into a symbol of national identity, while Bangla was pushed aside.¹²⁵ Dissent was not just political but almost sinful.

India entered the conflict with its own symbolic framing. Faced with millions of refugees crossing into its borders, it cast itself as a humanitarian protector. Indian leaders described their intervention as an act of justice, standing up for democracy against oppression. On a deeper level, India used the crisis to challenge Pakistan's founding myth, presenting the break-up of East Pakistan as proof that the "two-nation theory" had failed.¹²⁶ For India, the birth of Bangladesh was more than a geopolitical victory; it was a symbolic correction of history.

Eventually, Bangladesh gained independence after a 13-day conflict with the Indian army. As the war ended, a new nation was born, creating a national identity that represented all Bengalis of the region, free from the oppression they had faced.

Jammu and Kashmir: The disputed land of Jammu and Kashmir

In the conflict over Jammu and Kashmir, symbols carry a weight far beyond their surface meaning, and flags are one of the clearest examples. For India, the tricolor represents sovereignty and the belief that Kashmir is an inseparable part of the nation. Pro-India groups and government forces raise it to make that point visible. On the other side, pro-Pakistan and separatist groups often wave the green Pakistani flag as a way of resisting India's authority and expressing a demand for self-determination or even accession to Pakistan. For decades, the red Jammu and Kashmir state flag with its plough symbolized a sense of regional identity and autonomy under Article 370. When Article 370 was revoked in 2019, and that flag was removed, many Kashmiris felt a deep sense of loss, as if part of their cultural identity had been erased.¹²⁷ Because of this, flags aren't just cloth on a pole in Kashmir. They're powerful statements that turn rallies, protests, and even funerals into stages for political expression.

Religion also plays a major role in shaping how people come together around the conflict. Mosques and shrines often become spaces where political and spiritual life intersect. For example, Friday prayers frequently end with chants for "Azadi," or freedom, transforming what would normally be a religious gathering into a political demonstration.¹²⁸ These moments show how easily sacred spaces take on political meaning, and how the conflict itself is often framed not only as a struggle for land or autonomy, but as something deeper, almost spiritual. By blending faith and politics, religious symbols give the struggle moral weight, making many people feel that their cause is not just a political demand but also a matter of justice and destiny.

In the end, symbolism has been a powerful way for people in Jammu and Kashmir to express their feelings and struggles. Flags, prayers, landscapes, and even everyday culture have all taken on meanings far beyond what they normally signify. A flag isn't just a piece of cloth, a mountain isn't just a view, and a prayer isn't just a ritual; they become ways to show identity, resistance, and hope. These symbols help people share their story and keep their cause alive, which is why they remain such an important part of the conflict.

¹²⁵ Srinath Raghavan, *1971: A Global History* (Cambridge: Harvard University Press, 2013). 213-216

¹²⁶ Bose and Jalal, *Modern South Asia*. 203-205

¹²⁷ Sumantra Bose, *Kashmir: Roots of Conflict* (Cambridge: Harvard University Press, 2003). 203-204

¹²⁸ Upadhyay, *Secular Democracies*. 141-144

Sri Lanka

Sri Lanka's independence in 1948 was not achieved through a sudden revolution. Rather, it occurred through a gradual struggle shaped by ideas, negotiations, and the use of powerful symbols that brought people together. Under British rule, Sri Lankans came to see the importance of shared imagery drawn from religion, culture, and national identity as a way to inspire hope and unite different communities in the fight for self-rule.

One of the most important symbols was the lion, an ancient emblem of Sinhalese kingship that later became central to the national flag.¹²⁹ Its revival during the independence movement reminded people of a proud past before colonization and strengthened the belief that freedom could be restored. Religious symbols also carried great weight. The Buddhist flag, first created in the late 19th century, became more than a symbol of faith.¹³⁰ It became a marker of cultural pride and proof that Sri Lankan traditions would endure despite foreign influence.

The movement also drew meaning from language and clothing. Using Sinhala and Tamil in public speeches instead of English emphasized the right to speak in one's own voice. Leaders who chose to wear traditional dress rather than Western suits demonstrated they stood with ordinary citizens rather than colonial elites. Even public gatherings at temples or beneath the Bodhi tree became rich with symbolism, transforming familiar places into spaces of strength, resilience, and unity.

Symbolism and National Identity in Post Soviet States

With the fall of the Soviet Union, Central Asian states became laboratories for illegitimate democracies and stalled progress. Countries like Kyrgyzstan, Uzbekistan, and Turkmenistan were in disarray, and their former communist leaders' struggles to maintain power led to the birth of new nations under fabricated pretenses. The political elite created traditions and changed culture to control and/or divide the people of their countries, furthering authoritarian control. I argue that political leaders created national myths and identities to unite or divide the common people and maintain control after the Soviet Union's fall.

Kyrgyzstan

After the dissolution of the Soviet Union, Kyrgyzstan, one of the most ethnically and culturally diverse Central Asian states, experienced a period of political uncertainty. Lacking a strong tie with Pre-Soviet traditions or robust institutions, the new President, Askar Akayev, faced a legitimacy vacuum. Without a unifying political ideology or economic stability, the country faced turmoil and disunity. So in an effort to change this, Akayev turned to cultural nationalism and national myth-making to promote unity and authority.¹³¹

Akayev's efforts in promoting national unity included the Manas Epic. A vast oral tradition recounting the heroic deeds of the warrior Manas and his descendants occupies a central place in Kyrgyz cultural memory. Following independence, President Akayev strategically revived this narrative as a foundation for the new nation's identity. Under his leadership, *Manas*

¹²⁹ K. M. de Silva, *A History of Sri Lanka* (Berkeley: University of California Press, 1981). 470-472

¹³⁰ K. M. de Silva, *A History of Sri Lanka* (Berkeley: University of California Press, 1981). 470-474

¹³¹ James G. Mellon, "Myth, Legitimacy and Nationalism in Central Asia," *Ethnopolitics* 9, no. 2 (2010): 188-190.

was transformed from a folk epic into a state-sponsored symbol of unity, endurance, and moral virtue, meant to bind a population fragmented by regional, ethnic, and linguistic divisions.¹³² The 1995 celebration of the *1,000th anniversary of the Epic of Manas* was emblematic of this project. It was not only a cultural commemoration but a political performance of nationhood, positioning Kyrgyzstan within a mythic continuum that stretched deep into a pre-Soviet past. Through the building of statues, the rewriting of school textbooks, and the inclusion of *Manas* in official rhetoric, Akayev sought to craft a sense of historical depth and legitimacy for the fledgling state. This deliberate myth-making can be understood as part of a broader post-Soviet strategy of invented tradition, in which symbolic heritage fills the void left by the collapse of communist ideology. In Akayev's Kyrgyzstan, *Manas* served both as a moral guide for citizens and as a legitimizing mirror for state authority, a narrative that fused cultural pride with political necessity.

More specifically, Akayev's appropriation of the *Manas* epic involved selective reinterpretation rather than faithful preservation. Traditionally, the epic existed in multiple versions, transmitted orally by *manaschys* who adapted the story across regions and generations. Under state patronage, however, this plurality was narrowed. Official narratives emphasized themes of obedience, unity under a single leader, and territorial defense, while downplaying elements of internal conflict, dissent, and negotiation that also exist within the epic tradition. In doing so, the state effectively froze a living cultural text into a rigid political instrument. The figure of Manas was reframed less as a contested heroic symbol shaped by communal memory and more as an emblem of centralized authority aligned with Akayev's presidency. This transformation marked a subtle but significant misappropriation: a communal narrative rooted in oral dynamism was converted into a top-down ideology, stripping it of its ambiguity and using cultural authenticity to legitimize contemporary power structures rather than to reflect historical reality.

Yet, as scholars such as Mellon (2009) suggest, the power of such mythic nationalism is inherently unstable: it can inspire unity, but it also risks masking deeper structural weaknesses.¹³³ The *Manas* revival thus reflects the broader tension between symbolic nation-building and substantive governance, a dynamic that lies at the heart of post-independence identity formation across Central Asia.

The deployment of the myths also filled several related functions. At the simplest level, it provided the political legitimation mechanism to permit Akayev to shape his leadership within the confines of a historically established past infused with heroism and moral authority. By framing himself in the shadow of the legendary hero Manas, the ultimate defender and unifier of the Kyrgyz people, Akayev could metaphorically place his presidency within an epic scope of history rather than an ephemeral post-Soviet venture. Furthermore, by elevating the stature of Manas, Kyrgyzstan distinguished itself from its larger and more dominant neighbors, Kazakhstan and Uzbekistan, by creating an indigenous and distinct cultural narrative grounded in an oral tradition rather than the Soviet one. The insistence on cultural distinctness became essential to

¹³² Mellon, "Myth, Legitimacy and Nationalism," 191–194.

¹³³ James G. Mellon, "Myth, Legitimacy and Nationalism in Central Asia," *Ethnopolitics* 9, no. 2 (2010) 201–203

the formation of an image of national cohesion to be transmitted to either the home audience or the broader international stage.

Nevertheless, reliance on symbolic nationalism revealed the Kyrgyz state's vulnerabilities. The unifying rhetoric embodied in *Manas* frequently conflicted with the harsh realities characterized by profound social and regional divides, most prominently the historical schism between the more urbanized and politically influential north and the economically disadvantaged south. These internal divisions were further exacerbated by ethnic frictions between Kyrgyz and Uzbek communities, along with the increasing disparities between rural and urban populations. In a broken country, there is no progress or improvement in the lives of its citizens. Without unity, there is no shared sense of responsibility, nor an identity that defines a culture. Ergo, there is a necessary unit needed amongst the people to promote and foster the improvement of a country. From false flag events to fabricated stories to incite patriotic bluster, governments across the world have used fake media and propaganda for political control. Through the analysis of Kyrgyzstan, I argue that this phenomenon not only occurred but also led to a civil war in the push for national unification.

In this regard, the *Manas* narrative functioned as both a unifying myth and a political façade, an idealized portrayal of harmony that obscured ongoing inequalities and ineffective governance. As Mellon and other scholars have noted, this phenomenon mirrors a broader post-Soviet tendency in which states exchange symbolic unity for institutional solidity, leveraging history and heritage to maintain authority amidst weak political underpinnings.

Uzbekistan

In Uzbekistan, the post-independence development of national mythology adhered to a comparable rationale of political legitimization; however, it adopted a more overtly authoritarian and grandiloquent expression under the leadership of President Islam Karimov. The government reconceptualized historical narratives to focus on Amir Timur (Tamerlane), the fourteenth-century conqueror whose dominion once extended over significant portions of Central Asia and beyond. Karimov positioned Timur as a national icon and moral model, portraying him as an emblem of strength, unity, and discipline, attributes that the regime asserted it exemplified in guiding the contemporary Uzbek nation. Through the establishment of statues, museums, textbooks, and an abundance of commemorative events, Timur's representation became a representation of an uninterrupted Uzbek civilization that existed before Russian and Soviet dominance.

While myth-formation was premised on a return to a dynamic and orally transmitted tradition, the Uzbek state's appeal to Timur was fixed on permanency, monumentalism, and discipline.¹³⁴ The Karimov leadership brought Timur's legacy to bear directly on space, from naming streets to overhauling the center of cities and insisting on Samarkand as the symbolic capital of the empire. The spatialized remembering of history was itself a methodological form of politicized education: Timur became a presence that was both immediate and unchallenged

¹³⁴ James G. Mellon, "Myth, Legitimacy and Nationalism in Central Asia," *Ethnopolitics* 9, no. 2 (2010) 192-194.

throughout society.¹³⁵ Thus, while remembering the past, society was also governed by it. Historiographical interpretations that might otherwise emphasize Timur as a typification of violence, disunity, or coercive imperialism were thereby subjected to the periphery, while his conquests were seen as achievements in the march toward civilization. This made history absolute as moral truth, with majesty and order as anxieties that naturalized state control over society as somehow inevitable. Instead of a call to collective memory, there was a deliberate state-managed respect for history that made history rather a silent teacher against the seditious implications of deviation from a magnified past that was presumably unified.

Fulfilling different political purposes, this resuscitation of historical accounts sought to legitimize Karimov's authoritarian leadership by connecting his rule to the reputation of an earlier period characterized by the imposition of order and the success of conquests. At the same time, it also allowed Uzbekistan to define its identity distinct from its Soviet past and its post-Soviet peers, thereby representing the state as the rightful heir to the region's former empire and cultural glory. Nevertheless, this unifying narrative of continuity and power veiled important contradictions. Celebrating the existence of a homogeneous past and its hero concealed ethnic diversity and regional complexity, thereby providing a homogenized version of Uzbek identity that excluded the non-Uzbek minority and also different versions of the past. According to Mellon (2009), this kind of mythic nationalism provided time-bound legitimacy but did not address the underlying shortage of governance, transparency, and inclusiveness that still endangered the state's political life.¹³⁶ This way, the cult around Timur also followed Karimov's broader strategy: the use of selective memory and cultural performance to entrench power even as actual political reform has not materialized.

Turkmenistan

In Turkmenistan, myth-making took its most extreme and dramatic form under the leadership of Saparmurat Niyazov, who adopted the appellation Turkmenbashi, "Father of the Turkmen." Unlike the historical borrowing of heroes common to its neighbors, Niyazov put himself at the center of the national myth and elevated leadership to the level of a quasi-religious story about destiny and devotion. His personality cult insinuated itself into all aspects of everyday life: Niyazov statues spun to greet the sun in the capital city, months of the year were named after him and his family members, and his book, the *Ruhnama*, became the sacred national book studied in schools and workplaces and required reading during the driver's tests.¹³⁷

The *simvolizatsia* under Niyazov's leadership differed not only by its degree of passion, but also, more profoundly, by its level of engagement.¹³⁸ The *Ruhnama*, far from being treated as allegorical texts or holy writ, became institutionalized as an ethical and metaphysical handbook

¹³⁵ James G. Mellon, "Myth, Legitimacy and Nationalism in Central Asia," *Ethnopolitics* 9, no. 2 (2010) 192-194.

¹³⁶ James G. Mellon, "Myth, Legitimacy and Nationalism in Central Asia," *Ethnopolitics* 9, no. 2 (2010) 193-195

¹³⁷ James G. Mellon, "Myth, Legitimacy and Nationalism in Central Asia," *Ethnopolitics* 9, no. 2 (2010) 201-202

¹³⁸ James G. Mellon, "Myth, Legitimacy and Nationalism in Central Asia," *Ethnopolitics* 9, no. 2 (2010) 200-201

on Turkmenitude, oscillating as it did between civic obligation and religious belief. The imprimatur of loyalty, inscribed variously into such prosaic routines of life as education, employment, and licensing, turned mythology on its head, causing it to serve as an instrument of socialization. The politicization of chronology suffused personal memory with ideological necessity, reducing personal experience to the symbolic footprint of the ruler. Thus, rather than being something to be born with or something contested, nationality became something practiced, something performed. The misappropriation, here, consists specifically in this transference of culture to duty, where, more than simple political opposition, any resistance became either heresy or blasphemy against an invented nationality.

The habit of self-mythologizing filled two main roles. At the national level, it sought to erase the distinction between the state, the nation, and the leader, and made loyalty to Niyazov synonymous with genuine patriotism. Internationally, it posited Turkmenistan as an integrated, steady, and distinctively unique culture that resisted outside influences. But behind this semblance of unity lay an ultra-centralized and repressive political system characterized by censorship, isolationism, and stagnation. Niyazov's successor, Gurbanguly Berdimukhamedov, initially attempted to distance himself from the Turkmenbashy period's excesses but subsequently restored similar signs of a personality cult, thereby continuing the trend towards personalized nationalism to fill the void in institutional legitimacy.

For Turkmenistan, myth functioned not merely as propaganda but as political theater, in which the lack of political participation was filled by spectacle and the continuum of the past was reinvented to legitimize autocracy.¹³⁹ Like Uzbekistan and Kyrgyzstan, the state's exploitation of cultural icons further revealed an underlying contradiction: the same myths that purported to unify and enshrine the nation also concealed, and sustained, the weakness of its political basis.

Tajikistan

In Tajikistan, the development of national identity following independence was significantly complicated by the repercussions of the civil war (1992–1997), which revealed profound political, regional, and ideological divides within the nation. In response to this turmoil, President Emomali Rahmon looked to the distant past to construct a unifying narrative that could mitigate the fractures caused by conflict. A central component of this nation-building initiative was the promotion of Ismaili Somoni¹⁴⁰, the founder of the Samanid Empire in the ninth century, as the symbolic forebear of the Tajik nation. By designating the Somoni period as the golden age of Tajik civilization, Rahmon aimed to establish a connection between contemporary Tajikistan and a Persianate, pre-Islamic heritage that set it apart from both its Turkic neighbors and the Soviet past. The establishment of the Somoni as the national currency, the renaming of streets and monuments, and the erection of the monumental Somoni statue in Dushanbe exemplified this endeavor to ground state legitimacy in a mythic historical narrative characterized by cultural sophistication and imperial splendor.

¹³⁹ James G. Mellon, "Myth, Legitimacy and Nationalism in Central Asia," *Ethnopolitics* 9, no. 2 (2010) 203-204

¹⁴⁰ James G. Mellon, "Myth, Legitimacy and Nationalism in Central Asia," *Ethnopolitics* 9, no. 2 (2010) 188-189

In terms of Tajikistan, the use of the Somoni¹⁴¹ also played a role in constructing a new system of governing memories after the civil war. As the regime elevated an obscure and historical figure that was more than 1000 years old (who did not come to great public attention until the civil war), there was little discussion about the traumatic open issues left behind by the conflict of the 1990s, including questions about accountability, reconciliation, and inclusion in the political process for various factions. By providing a new time reference for how to unify the country, this made the Somoni narrative a temporal means of establishing a new type of national unity. Through the paradigm of 'peace' as a return to the historical destination for the 'people', it created a framework for peace being a result of a return to an idealized time period, rather than a negotiated settlement between oppositional parties from the past. The notion of peace as returning to a higher state of being established peace as a myth and established legitimacy for current authorities; while the actual events and feelings of the war, such as regional issues, ideological divides, and loss of the social framework, became secondary to the depiction of the continued civilization of the people of Tajikistan. The problematic nature of the narrative is not related to celebrating a particular heritage; rather, it is not being used to construct the foundation for an honest political reconciliation process in an already fragmented society that continues to suffer the effects of civil war.

The historical resuscitation also met several strategic aims. At one level, it recast Tajikistan's postwar recovery as a civilizational renaissance, enabling Rahmon to portray himself as the author of national pride and unity. By evoking Somoni, the regime furthered an ideology of continuity and endurance that disguised the simultaneous processes of authoritarian entrenchment and economic vulnerability. At another level, this mythological construction was used to address ethnic and regional diversity between the northern elites of Leninabad (now Sughd) and the southern provinces, depleted by the civil war. Appeal to the shared Persianate culture allowed the state to define Tajik identity in cultural rather than purely ethnic terms and thus legitimize centralized authority under the guise of national unity.

Nonetheless, Mellon (2009) and other scholars note that the symbolic meaning of the Somoni myth was insufficient to overcome Tajikistan's ultimate crisis of legitimacy.¹⁴² While the imagery representing unity and temporal depth implied stability, it also concealed persistent problems of corruption, oppression, and economic dependence that still eroded the social contract between the state and its citizens. Here, Tajikistan's recourse to cultural revivalism also replicated the wider regional pattern: mythic nationalism became the fount of national pride and the means of control, acting in place of democratic legitimacy in a still war-ravaged and authoritarian state.

Conclusion

Throughout the twentieth century, warfare and revolution have repeatedly demonstrated that the most effective instruments of political and ideological power lie in the realm of symbols. In fascist Europe, in the aftermath of colonialism in South Asia, and in the newly formed

¹⁴¹ James G. Mellon, "Myth, Legitimacy and Nationalism in Central Asia," *Ethnopolitics* 9, no. 2 (2010) 193-194

¹⁴² James G. Mellon, "Myth, Legitimacy and Nationalism in Central Asia," *Ethnopolitics* 9, no. 2 (2010) 194-195

republics of the former Soviet Union, flags, mythologies, rites of passage, and other elements of culture have been employed to convert collective identity into an instrument of control-

In fascist Germany and Italy, symbols like the swastika and fasces were twisted into tools of control, blurring belief and obedience to enforce domination. In South Asia, symbols once used by leaders like Gandhi and Bhagat Singh to unite people against imperialism were later repurposed by new nations to divide identity through nationalism. Even after empires fell, leaders in former Soviet republics revived ancient heroes and legends to manufacture unity and legitimize power rather than promote true democracy.

Taken together, these examples from the above cases affirm the observation that symbols are neither purely good nor purely evil; their significance always depends on who uses them and for what ends. Whether embroidered on a flag, inscribed on a monument, or printed on fabric, symbols are always tools of politics capable of stirring a sense of common purpose and obscuring the power of oppression alike. Through the upheavals of revolution and warfare in the last century, the symbols of nations are seen to unite as well as manipulate to mobilize public opinion, demonstrate power, and command compliance in the name of membership in the group.

In contemporary American politics, symbols are often used to flatten political complexity into easily recognizable markers of allegiance. Objects such as the red “Make America Great Again” cap or even the American flag itself are frequently repurposed at rallies and protests to signal loyalty to a specific political movement rather than to shared civic values. This symbolic strategy became especially visible during the January 6 Capitol riot, when Confederate flags and “1776” imagery were carried through the Capitol to frame violence as patriotic rebellion. These symbols were not incidental; they functioned to legitimize unlawful action by invoking a selectively constructed national past. Similar dynamics appear in the repeated invocation of “law and order” through campaign banners and police imagery, where promises of security mask the expansion of force and surveillance. Informed by the historical patterns examined in this research, readers should approach such symbols with caution, asking not what they claim to represent, but how they are being used to demand compliance, exclude dissent, and redefine American identity in the present.

Works Cited

- Atkins, Jennifer. *Wearing Propaganda: Textiles on the Home Front in Japan, Great Britain, and America during World War II*. New Haven: Yale University Press, 2005.
- Bose, Sugata, and Ayesha Jalal. *Modern South Asia: History, Culture, Political Economy*. 4th ed. London: Routledge, 2017.
- Bose, Sumantra. *Kashmir: Roots of Conflict, Paths to Peace*. Cambridge: Harvard University Press, 2003.
- Chatterjee, Partha. *The Nation and Its Fragments: Colonial and Postcolonial Histories*. Princeton, NJ: Princeton University Press, 1993.
- de Silva, K. M. *A History of Sri Lanka*. Berkeley: University of California Press, 1981.
- Dunnage, John. "Building Ideological Bridges: The Police and the National Community in Italy, 1919–1943." *Crime, History & Societies* 19, no. 1 (2015): 85–104.
- Dunnage, John, and Nadine Rossol. "Building Ideological Bridges and Inventing Institutional Traditions." *Crime, History & Societies* 19, no. 1 (2015): 85–104.
- Falasca-Zamponi, Simonetta. *The Aesthetics of Politics: Symbol, Power and Narrative in Mussolini's Fascist Italy*. London: SAGE Publications, 1992.
- Guha, Ramachandra. *India After Gandhi: The History of the World's Largest Democracy*. New York: HarperCollins, 2007.
- Khan, Yasmin. *The Great Partition: The Making of India and Pakistan*. New Haven: Yale University Press, 2007.
- Mellon, James G. "Myth, Legitimacy and Nationalism in Central Asia." *Ethnopolitics* 9, no. 2 (2010): 187–204.
- Raghavan, Srinath. *1971: A Global History of the Creation of Bangladesh*. Cambridge, MA: Harvard University Press, 2013.
- Upadhyay, Anuradha. *Secular Democracies, Governance, and Politics of Religion in South Asia*. New Delhi: Routledge, 2018.

How Do The Actual Economic Impacts Of the Super Bowl and Coachella Compare To The Implications Set by the Companies and the Media? By Ruchika Varanasi

Abstract

The Super Bowl and Coachella are known for being huge events, bringing in millions of dollars every year to the Coachella Valley and the host city of the Super Bowl. These events enhance tourism by encouraging attendees to eat at restaurants and stay at nearby hotels. However, in recent research, the Super Bowl and Coachella's estimated financial effects have been shown to be inflated in comparison to the actual results. Using research papers and news articles, I aim to analyze how the actual economic impacts of the Super Bowl and Coachella compare to the narratives portrayed by the companies and media.

Introduction

The United States is known all around the world for being the hub of entertainment and sports; with its many concerts and games, there are many ways to see the best in music and sports. However, some events are more popular than others, so popular, in fact, that people will spend hundreds of dollars on plane tickets and hotels just to make sure they can attend, along with thousands of dollars for tickets. Two of these events are called Coachella and the Super Bowl, and they are the most admired events in the sporting and music industry, with thousands of people attending them every year. While other global events, such as the World Cup, also take place in the US at times, the Super Bowl and Coachella are solely the US's heart and pride.

The Super Bowl started in 1967, due to a merger between the NFL and the AFL, and it was broadcast on live television for many to see. It was nowhere near as popular as today, however, as the crowd in the stadium was "less than sellout" (Carroll). The name 'Super Bowl' and the use of roman numerals were eventually introduced in 1969 (Carroll). Nowadays, it's extremely popular as it is an unofficial national holiday of sorts, with people looking eagerly to see who will win and making bets on it, spending money on takeout food, and hosting parties with loved ones. On the corporate side, companies are spending millions of dollars on 30-second advertisements, trying to acquire new customers in the limited time they're allotted, or even signing up to be official sponsors of the Super Bowl. For example, Apple Music has been an official sponsor for the past 3 years. In the process of being a sponsor, they have acquired 5 million new subscribers, going from 88 million in 2022 to 93 million in 2025 (<https://www.facebook.com/kulwant.nagi>). This shows that their brand has gotten a lot of awareness and a lot more publicity due to this sponsorship, which, in turn, increases their revenue. Often, these advertisements and the halftime show are considered the best part of the Super Bowl, appealing to even those who don't understand the game.

In the music industry, Coachella is known for having some of the biggest pop stars come and perform. It initially started in 1999 and has only grown bigger since, even sparking inspiration for copycat festivals like Fyre. Artists such as Lady Gaga, Post Malone, and Travis Scott headlined the 2025 festival, but additional artists like Benson Boone and Ed Sheeran also performed. Many people are attracted to Coachella as it gives them the opportunity to see many of their favorite artists without spending money on concert tickets for each artist, and also to

meet influencers who attend these events. It also enhances tourism to the Coachella Valley, bringing attractions to small businesses in the area, which in turn boosts the local economy. In contrast to the Super Bowl, Coachella does the majority of its advertising through influencers on social media, paying them thousands of dollars for one post.

Some main differences between Coachella and the Super Bowl are that Coachella takes place over three days for two consecutive weekends in April in order to accommodate the many people who want to attend, and the Super Bowl takes place on one Sunday in February. Additionally, the Super Bowl takes place in different cities, its location changing each year, but mainly in the big stadiums such as Levi's Stadium in Santa Clara and SoFi Stadium in Los Angeles. In contrast, Coachella takes place in the Coachella Valley each year, hence the name "Coachella." In terms of ticket pricing, Coachella tickets range from around \$589 to \$1,299 for weekend 1 and \$539 to \$1,139 for weekend 2 (Coachella 2020 | Pass Types), whereas the Super Bowl spans \$3,000 to \$12,000 on Ticketmaster (Czachor). As a result, according to the media, the overall projected profits for the Super Bowl and Coachella in 2025 were \$1 billion and \$700 million, respectively. Finally, Coachella is owned by GoldenVoice, and the Super Bowl is controlled by the National Football League (NFL).

Over the years, the media has stated that the economic impact of these events has been huge, leading many cities to compete to try to host the Super Bowl. Additionally, Coachella has become increasingly popular in the news with its significant impact on the local economy. However, recent research has come to light stating that the Super Bowl numbers have been vastly inflated and the actual economic impacts are only $\frac{1}{4}$ of the projected amount the NFL states, whereas Coachella's numbers have also been inflated, but to the extent it is still being discovered. Nevertheless, using research papers and news articles, I aim to analyze how the actual economic impacts of the Super Bowl and Coachella compare to the narratives portrayed by the companies and media that cover these events. Some of the indicators I will use are the number of attendees, hotel or Airbnb reservations, the total profit, and tourism activity. As part of my literature review, I will use specific examples of recent Super Bowls and Coachella festivals in the last 5-10 years.

Methodology

My methodology is that I will analyze news articles and research papers from various sources, such as the US Chamber of Commerce, WeRave.com, HolyCross, NBC, and Digital Commons. I will use 10-15 research papers and 3-5 news articles as part of my research. I have 11 sources for the Super Bowl and 9 sources for Coachella. To ensure an unbiased outlook, I chose these sources because they provide different perspectives and information on these topics.

Literature Review

This section analyzes the overall state of the US economy during the 2025 Coachella and Super Bowl dates, then looks at Coachella and Super Bowl's individual economic impact both in 2025 and years past, and lastly explores their similarities and differences.

This section analyzes the state of the US economy during February and April 2025, which is when the Super Bowl and Coachella were held, respectively. In February, Consumer

Confidence decreased from January 2025, falling to 98.3, a steep drop of 7 points. This type of steep drop had not happened since August 2021. The Expectations Index (the short-term outlook of market conditions) of Consumer Confidence dropped to 72.9, indicating a recession ahead since it was below 80. The unemployment rate stayed the same at 4.1%, but the Federal Labor Department added 151,000 new jobs, increasing the workforce. Additionally, GDP decreased by 0.5% during the first quarter of 2025 (January, February, and March). These numbers were all influenced by current politics, such as Trump's threat of tariffs, inflation rates (2.8%), and volatility in the technology sector, with concerns about new advancements in AI, such as companies like DeepSeek emerging. Overall, February was showing more positive trends, but still weaker than the final quarter of 2024.

In April 2025, which is when Coachella was held, Consumer Confidence hit its lowest since August 2020 with a score of 86.0, and Expectations Index hit its lowest since October 2011 with a score of 54.4, indicating a major recession. Additionally, the unemployment rate remained steady at 4.2%, with 177,000 jobs added. GDP fell by 0.3% after a 0.2% increase in March, but still showed a total 0.7% increase from January to April. All of this was due to President Donald Trump announcing his tariffs on various countries, also known as "Liberation Day" on April 2nd. As a result, retaliatory tariffs were announced from places such as China, and also contributed to the influence on stocks and these numbers. Fear from consumers also led to this crash. The inflation rate was 2.3%, which was a decrease from February 2025.

In terms of individual markets, the US music tourism market was estimated at 27.58 billion in 2024 and expected to grow at a CAGR (Compound Annual Growth Rate) of 15.5% from 2025 to 2033, according to GrandViewResearch. As for the US sports market, it was estimated at an annual revenue of 150.0 million in 2024, and it's expected to grow to 208 million by 2030 with a 5.6% CAGR, as stated by GrandViewResearch. These numbers show a sign of positive economic growth for these two industries, which is important because it indicates that these industries will continue to flourish for a long time. Additionally, arts and cultural economic activity accounts for 4.2% of the GDP, equaling around 1.17 trillion, according to Statista. Their percentages of profits in comparison to the GDP, however, are too small to compare.

The Super Bowl in 2025 was held in New Orleans, and according to the Louisiana Economic Development Organization (a state organization that helps support businesses, similar to the US Chamber of Commerce), it had a "positive economic effect on the Louisiana economy," generating an estimated \$1.25 billion and over 100,000 visitors from out of state, totaling around 115,000 attendees for the actual event. Around 9,800 jobs were created solely for this event, and \$395 million in wages were given. Coachella Valley's economic impact, according to NBC Palm Springs, was around \$704 million, and ticket revenue was strong, at around \$128 million. The PSP (Palm Springs International Airport's Operations) created around 18,000 jobs during that month. According to the Coachella Valley Chamber of Commerce, it included jobs from festival tourism, and it's likely that those numbers went down after the festival since those jobs were no longer needed.

Super Bowl

The US Chamber of Commerce is a government organization that supports businesses, both large and small. They also have a partnership with the NFL called the “Brought to You by American Business” campaign, which showcases the positive economic impacts the Super Bowl has on a host city. According to the US Chamber of Commerce, the projected profit for the 2025 Super Bowl, held in New Orleans, was estimated to be about \$500 million, but “the final figure could be much higher.” They also projected 125,000 visitors and \$200 million in spending, with an additional \$53 million in spending from natives in the area. However, the Louisiana Economic Development stated these numbers ended up being significantly higher than the numbers previously projected above. But these numbers do not match Statista, since according to them, the total number of visitors for the 2025 Super Bowl ended up being only around 65,000, as shown by Figure 1. Statista is an online platform that specializes in data gathering and visualization and uses information from market reports, online research, and more. It also goes through a rigorous peer review process when publishing this data.

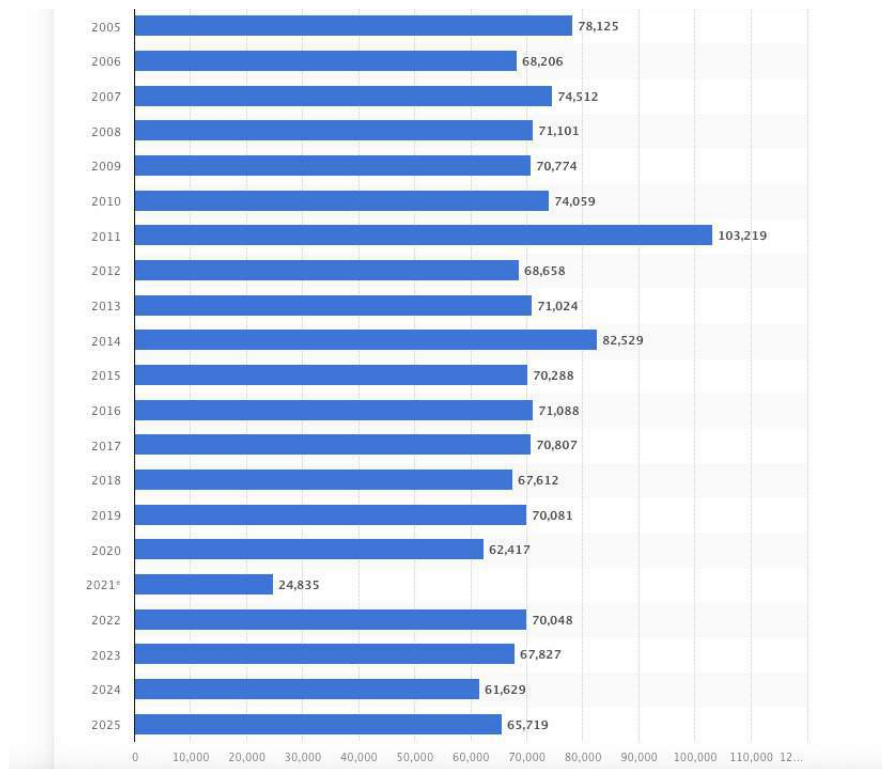


Figure 1: Attendance at the Super Bowl from 2005-2025 - www.statista.com/statistics/1362951/super-bowl-attendance/

However, this could indicate bias as the NFL wants cities to host the Super Bowl, and since both the US Chamber of Commerce and Louisiana Economic Development partner with them, they can influence the numbers in their favor. This also makes sense why these numbers ended up being significantly higher than other research papers have stated. For example, for the 2001 Super Bowl, Victor Matheson and Robert Baade from their study, “Assessing the Economic Impact of the Super Bowl,” demonstrated the notion that host cities couldn’t make more than \$300 million, since other costs significantly consume their profits. From 1995 to 2003, \$6.4

billion would have been spent cleaning up these stadiums and ensuring they're ready for hosting, which is a significant number considering most of this funding comes from state and local governments, where the money could be used elsewhere. This is because the NFL has certain requirements for its stadiums to host. Other states like Arkansas and Georgia were estimated to have millions of dollars in spending as well, ranging from 11 million to upwards of 42 million in spending, which is not corroborated by what the US Chamber of Commerce stated.

On the other hand, the Super Bowl creates lots of new job opportunities. As mentioned above, the 2025 Super Bowl created 9,800 jobs, according to Louisiana Economic Development. These numbers are corroborated by a research paper by Jason Luo, who found that between 2010-2022, when the data was collected, the unemployment rate for host cities is lowest in the time of 3 months before the event, with the greatest increase in the labor market happening 1 month before the event. The unemployment rate then becomes the highest 1 month after the event, before returning to its normal state in about 3 months. This corroborates what we've seen with the number of jobs created in the area, as you can see by Table 1, with January 2025 having the highest amount of jobs and a decrease in employment afterwards for 2 months before returning to normal levels.

Data Series	Back Data	Jan 2025	Feb 2025	Mar 2025	Apr 2025	May 2025	June 2025
Labor Force Data							
Civilian Labor Force(1)		2,077.4	2,073.4	2,071.1	2,069.4	2,071.2	^(p) 2,074.3
Employment(1)		1,983.4	1,981.3	1,979.5	1,978.4	1,978.8	^(p) 1,981.6
Unemployment(1)		94.0	92.0	91.5	91.0	92.4	^(p) 92.7
Unemployment Rate(2)		4.5	4.4	4.4	4.4	4.5	^(p) 4.5

Table 1: Louisiana Labor Force Data

In conclusion, the Super Bowl's economic impact has not been the same in expectation and reality, especially with the number of people attending the event and spending, but the labor market has been the same in reality and expectation.

Coachella

In 2025, as previously mentioned, the ticket revenue was estimated to be over \$100 million, since an estimated 250,000 visitors attended across both weekends, according to Matchfy. According to a research paper, however, from 2013 to 2015,, the estimated ticket sale revenue was around \$13 million. This could be due to a few reasons. For example, over time, inflation has significantly risen (1.3% in 2013 versus 2.7% in 2025), and only 25,000 people attended in 2013. The inflation rate difference leads to ticket prices being different because, to keep up with profits, they have to price tickets higher, which makes the overall profits seem higher. This is because Coachella was also not as popular back in 2013 as it is now, since there wasn't as much emphasis on its marketing and its performer lineup. In contrast to the Super Bowl, since Coachella is held in the same place every year (Coachella Valley in Indio), most of the refurbishments can be reused from year to year, significantly lowering costs for them in that aspect. However, costs from marketing and performers are insignificant; according to Yahoo

Entertainment, they're paid anywhere from \$600,000 to upwards of \$12 million, depending on the performer. Bigger performers, like the Weeknd or Ariana Grande, receive more money. Marketing costs are around the same, with around \$600,000 for one 30-second ad, but a lot of their advertising comes from influencers and also mentions and hashtags on pages like Instagram and TikTok from people who attend these festivals. According to Hammerhead Global, about $\frac{1}{3}$ of the money Golden Voice makes from Coachella gets injected back into Indio (around 36 million for 2023). These numbers are slightly different from NBC Palm Springs, which projected around a \$100 million profit in Indio, and even bigger in Matchfy, which estimated \$700 million.

The reason for these discrepancies is due to a few reasons. For one, these sources could be biased, since they may partner with Coachella directly to come up with these numbers, similar to the Super Bowl. Additionally, the performers who performed that year, the state of the local economy, and the inflation rates can all affect the overall profits, but not to this significant extent, to where there's a 600 million difference in profit from sources in one year. However, all the sources agree that when Coachella was not held during the pandemic, it caused a significant economic depression for Indio. Spending in Coachella for 2025 had significant impacts, according to Fox Business, as tickets, food, hotel stays, and flights end up being around \$3,000+ for a consumer. The exact spending numbers were not provided for 2025, but it's estimated to be in the millions if every consumer is spending around this much. Some sources corroborate these numbers (Julian Perez, 2017, and Statista, 2025).

For their job market, as previously mentioned, it was estimated to have a significant impact, creating 18,000 jobs.

Civilian Labor Force

Jun 24	Apr 25	May 25 Revised	Jun 25 Prelim	Percent Change Month	Percent Change Year
2,186,700	2,233,800	2,231,700	2,232,400	0.0%	2.1%

Table 2: Riverside-San Bernardino-Ontario MSA Press Release - labormarketinfo.edd.ca.gov/geography/riverside-county.html.

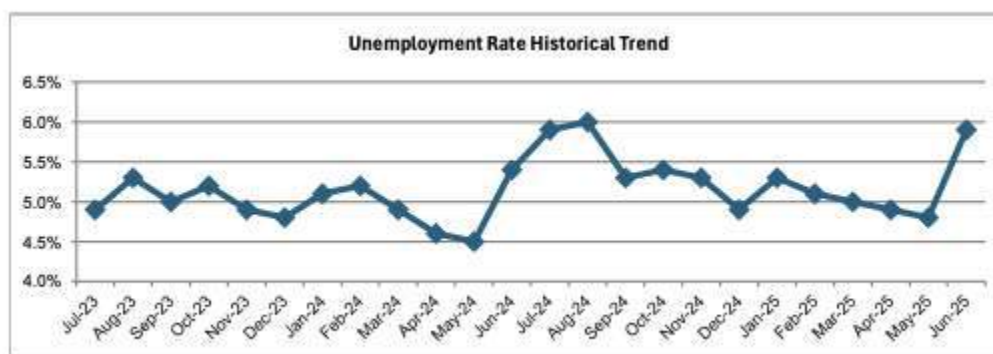


Table 3: Riverside-San Bernardino-Ontario MSA Press Release - <http://labormarketinfo.edd.ca.gov/geography/riverside-county.html>

2025	Labor Force	Employment	Unemployment	Unemp. Rate
January	1,194,000	1,130,200	63,800	5.3%
February	1,192,800	1,130,900	61,900	5.2%
March	1,201,800	1,140,700	61,100	5.1%
April	1,193,800	1,135,400	58,400	4.9%
May	1,193,000	1,134,800	58,200	4.9%
June	1,193,500	1,121,900	71,700	6.0%

Table 4: Riverside County Labor Force Data - labormarketinfo.edd.ca.gov/geography/riverside-county.html.

The data from Tables 2, 3, and 4 are from Riverside County, the county where Coachella is held. According to the US Department of Labor Statistics, the unemployment rate became its lowest during the months leading up to Coachella (February and March) and after Coachella (May) before shooting up again in June, indicating that the setup and clean up, the hotels, and all the other labor they need in those months employ a significant amount of people. This has been the trend for the past 2 years, as shown by Table 3. All of this says the same thing, which is that Riverside County had a significant labor market impact during the months of Coachella. In conclusion, Coachella's overall revenue numbers have discrepancies from source to source, but the labor market and ticket market numbers are around the same across sources.

Limitations

Some limitations encountered while creating this literature review were biases from various sources, especially governmental or news organizations. This is because many of these governmental organizations partner with the NFL to come up with these numbers, so they get their data from them, which, as we saw, ended up being inflated. There were not many research or academic papers available about Coachella, whereas the Super Bowl had a lot of information previously researched and available. The Coachella section mainly used blogs and news articles, which could mean bias towards the numbers and the information available since they don't come from research-driven methods, but rather, information that's already public. Finally, the economy at the time of this data analysis (February and April 2025 and past years) was influenced by other external factors, such as in April 2025 when tariffs and Liberation Day were announced, already bringing down numbers like Consumer Confidence and stocks. In February 2025, AI and volatility in stocks were a worry for many consumers as well. In the past 10 years, other events such as fluctuations in the stock market, rising inflation rates over this period, and other announcements during these times would affect the local economy as well. This is due to uncertainty among consumers, which leads fewer people to want to attend these events or pay money for tickets.

Recommendations

This research suggests that cities considering hosting the Super Bowl should examine all the numbers involved (number of attendees, labor, refurbishing, etc.), not just profit. For both events, a recommendation is to do more research on the overall profits and make sure that the data going into it is accessible for everyone. Some areas for further investigation are hospitality

rates in both cities, whether hotel rates increase because of these events, and how that affects numbers, as well as looking at more trends across longer periods of time. One final recommendation is to perform similar research for other events around the world and look into these indicators to see if they follow a similar pattern.

Conclusion

To compare the Super Bowl's and Coachella's overall expectation versus reality, we can see that both Coachella and Super Bowl have had the similar increase in jobs from media and research, showing a steady growth in jobs leading up to the event, a significant amount before and during, and a drop after to its previous levels, indicating the importance of the event for temporary employment. For Coachella, the ticket revenue ended up being around the same range, agreeing on about \$100-\$130 million in ticket revenue across both weekends. In contrast, the Super Bowl's spending and visitor numbers, which contribute to ticket revenue, differed significantly from research and statistics (from Statista and other sources). Finally, their overall profits were very different from the media, with Coachella numbers ranging a significant amount (\$30 million to \$700 million). Super Bowl's overall profits had discrepancies between the government organizations(US Chamber of Commerce) and the state departments(Louisiana Economic Development and Greater Coachella Valley Chamber of Commerce), but the research paper profit numbers (Jason Luo, Robert Baade, and Victor Matheson) were relatively the same throughout.

Works Cited

- Altus, Kristen. "California's Most Popular Music Festival Comes with a Side of 'Diabolical' Sticker Shock." Fox Business, 18 Apr. 2025,
www.foxbusiness.com/media/californias-most-popular-music-festival-comes-side-diabolical-sticker-shock. Accessed 10 Aug. 2025.
- Ashley, Lockwood. The Super Bowl's Economic Impact on Its Host City Super Bowl's Economic Impact on Its Host City. Apr. 2018,
http://digitalcommons.assumption.edu/cgi/viewcontent.cgi?article=1035&&context=honorstheses&&sei-redir=1&referer=https%253A%252F%252Fscholar.google.com%252Fscholar%253Fstart%253D10%2526q%253Dsuper%252Bbowl%252B%252Bimpact%252Bon%252Bhotels%252Band%252Brestaurants%2526hl%253Den%2526as_sdt%253D0%252C5#search=%22super%20bowl%20impact%20hotels%20restaurants%22. Accessed 10 Aug. 2025.
- Board, The Conference. "US Consumer Confidence Dropped Sharply in February." Prnewswire.com, Cision PR Newswire, 25 Feb. 2025,
www.prnewswire.com/news-releases/us-consumer-confidence-dropped-sharply-in-february-302384647.html. Accessed 10 Aug. 2025.
- Carroll, Bob. "Super Bowl | History & Results." Encyclopædia Britannica, 15 Mar. 2019,
www.britannica.com/sports/Super-Bowl. Accessed 10 Aug. 2025.
- "Coachella 2020 | Pass Types." Ww.coachella.com, www.coachella.com/passes.
- "Consumer Confidence Declined for Second Consecutive Month in February." The Conference Board, Feb. 2022,
www.conference-board.org/topics/consumer-confidence/Consumer-Confidence-Feb22. Accessed 10 Aug. 2025.
- "Consumer Prices up 2.3 Percent from April 2024 to April 2025." Bureau of Labor Statistics, 19 May 2025,
www.bls.gov/opub/ted/2025/consumer-prices-up-2-3-percent-from-april-2024-to-april-2025.htm. Accessed 10 Aug. 2025.
- Czachor, Emily Mae. "How Much Were Super Bowl 2025 Tickets? See Average, Cheapest, and Most Expensive Prices for the Game." Cbsnews.com, 6 Feb. 2025,
cbsnews.com/news/how-much-super-bowl-tickets-2025/. Accessed 10 Aug. 2025.
- Dubay, Curtis. "The Super Bowl's Economic Impact in Cities across the Country." Uschamber.com, 6 Feb. 2025,
www.uschamber.com/economy/how-the-super-bowl-creates-economic-impact-across-the-country. Accessed 10 Aug. 2025.
- Matheson, Victor, and Robert Baade. (PDF) Padding Required: Assessing the Economic Impact of the Super Bowl, Sept. 2004,
www.researchgate.net/publication/5092492_Padding_Required_Assessing_the_Economic_Impact_of_the_Super_Bowl. Accessed 10 Aug. 2025.
- "The Super Bowl's Economic Impact in Cities across the Country." Uschamber.com, 6 Feb. 2025,
www.uschamber.com/economy/how-the-super-bowl-creates-economic-impact-across-the-country. Accessed 10 Aug. 2025.

Economics, Season. "Festival Season Economics." The Judy Stakee Company, 13 May 2024, www.judystakee.com/blog/festival-season-economics.

Enrico Novazzi. "Coachella: A Cultural Colossus in Music, Trends and More." Matchfy.io Blog News, 17 Apr. 2025, blog.matchfy.io/coachella-2025-a-cultural-colossus-powering-music-trends-and-millions/. Accessed 10 Aug. 2025.

Evolve Editors. "Here's How Much Artists Get Paid to Perform at Coachella." Yahoo Entertainment, 14 Apr. 2025, www.yahoo.com/entertainment/articles/much-artists-paid-perform-coachella-111559174.html. Accessed 10 Aug. 2025.

Gough, Christina. "Super Bowl Attendance 2023." Statista, 13 Feb. 2024, www.statista.com/statistics/1362951/super-bowl-attendance/. Accessed 10 Aug. 2025.

Hise, Emmy. "Hotel Rates Unprecedented for Las Vegas' Super Bowl Weekend." CoStar, costar.com, 27 Feb. 2024, www.costar.com/article/1172846186/hotel-rates-unprecedented-for-las-vegas-super-bowl-weekend. Accessed 10 Aug. 2025.

<https://www.facebook.com/kulwant.nagi>. "Affiliate Booster." Affiliate Booster, 15 Aug. 2025, www.affiliatebooster.com/apple-music-statistics/. Accessed 31 Aug. 2025.

Jason, Luo. Host Cities' Labor Market Reactions to the Super Bowl. 10 Mar. 2023, scholar.archive.org/work/khwjd3si7ra2lgw3cbl4r5s23e/access/wayback/research-archive.org/index.php/rars/preprint/download/828/1280/1059. Accessed 10 Aug. 2025.

Julian, Perez. Music Festivals: A Secondary Market Analysis. 25 Apr. 2016, scholarship.claremont.edu/cgi/viewcontent.cgi?article=2465&&context=cmc_theses&&sei-redir=1&referer=https%253A%252F%252Fscholar.google.com%252Fscholar%253Fhl%253Den%2526as_sdt%253D0%25252C5%2526q%253DCoachella%252Bimpact%252Bon%252Bthe%252Bconomy%252Bfor%252Bhospitality%252Band%252Btourism%2526btnG%253D#search=%22Coachella%20impact%20economy%20hospitality%20tourism%22. Accessed 10 Aug. 2025.

Kale, Rahul. "Coachella 2023: Impact of Festival on the Local Economy." We Rave You, 7 Apr. 2023, weraveyou.com/2023/04/coachella-2023-impact-on-local-economy/. Accessed 10 Aug. 2025.

"Louisiana Economy at a Glance." Wwww.bls.gov, <https://bls.gov/eag/eag.la.htm>. Accessed 10 Aug. 2025.

Molinari, Nicholas. "Super Bowls Are an Annual American Economic Sports Event." Uschamber.com, 6 Feb. 2025, www.uschamber.com/economy/super-bowl-lix-how-american-business-scores-the-biggest-night-in-sports. Accessed 10 Aug. 2025.

Moore, Andrew. "Super Bowl: A Money Machine, but for Whom?" College of Natural Resources News, 5 Feb. 2024, cnr.ncsu.edu/news/2024/02/super-bowl-economic-impact/. Accessed 10 Aug. 2025.

National Endowment for the Arts. "Arts & Cultural Sector Hit All-Time High in 2022 Value Added to U.S. Economy." Wwww.arts.gov, 25 Mar. 2024, www.arts.gov/news/press-releases/2024/arts-cultural-sector-hit-all-time-high-2022-value-added-us-economy. Accessed 10 Aug. 2025.

- “PSP Generates \$2.5 Billion into Regional Economy, Supports 18,000+ Jobs.” Greater Coachella Valley Chamber of Commerce, 4 June 2025, gcvcc.org/psp-generates-2-5-billion-into-regional-economy-supports-18000-jobs/. Accessed 10 Aug. 2025.
- Raymundo, David. IT’S the ECONOMY, STUPID! SUPER BOWL IMPACT on LOCAL ECONOMY of PUBLICLY SUBSIDIZED NFL STADIUMS: A Project. 2015, scholarworks.calstate.edu/downloads/vx021h564. Accessed 5 Aug. 2025.
- Smart, Tim. “Economic Indicators Fell Further in February, Weighed down by Consumer Pessimism.” US News & World Report, U.S. News & World Report, 2025, www.usnews.com/news/economy/articles/2025-03-20/economic-indicators-fell-further-in-february-weighed-down-by-consumer-pessimism. Accessed 10 Aug. 2025.
- Springs, NBC Palm. “Coachella Festival Brings Major Sales Boost to Indio and Local Valley Businesses.” Nbcpalmsprings.com, NBC Palm Springs, 12 Apr. 2025, www.nbcpalmsprings.com/2025/06/03/coachella-festival-brings-major-sales-boost-to-indio-and-local-valley-businesses. Accessed 10 Aug. 2025.
- Statista. “Percentage Added to U.S. GDP 2018, by Industry | Statista.” Statista, Statista, 2018, www.statista.com/statistics/248004/percentage-added-to-the-us-gdp-by-industry/. Accessed 10 Aug. 2025.
- “Super Bowl LIX’s Stunning \$1.25 Billion Impact for Louisiana— Impacting 10,000 Jobs & a Huge Economic Development Win - LED | Louisiana Economic Development.” Opportunitylouisiana.gov, 3 June 2025, www.opportunitylouisiana.gov/news/super-bowl-lixs-stunning-1-25-billion-impact-for-louisiana-impacting-10000-jobs-a-huge-economic-development-win. Accessed 10 Aug. 2025.
- “The United States Sports Event Market Size & Outlook, 2030.” Grandviewresearch.com, 2025, www.grandviewresearch.com/horizon/outlook/sports-event-market/united-states. Accessed 10 Aug. 2025.
- “U.S. Music Tourism Market Size | Industry Report, 2033.” Grandviewresearch.com, 2024, www.grandviewresearch.com/industry-analysis/us-music-tourism-market-report. Accessed 10 Aug. 2025.
- “Unemployment Rates and Labor Force.” Ca.gov, 2023, labormarketinfo.edd.ca.gov/geography/riverside-county.html. Accessed 10 Aug. 2025.
- Wallace, Alicia. “The US Economy Added 151,000 Jobs in February in First Full Jobs Report under Trump’s Second Term.” CNN, 7 Mar. 2025, www.cnn.com/2025/03/07/economy/us-jobs-report-february-final. Accessed 10 Aug. 2025.
- “Why Coachella Music Festival Sets the Marketing Bar.” Hammerhead. Global, 2025, www.hammerhead.global/blogs/why-coachella-music-festival-sets-the-marketing-bar. Accessed 10 Aug. 2025.

How Do Airdrops Influence Long-Term User Retention in DeFi Ecosystems?

By Bassel El Hage

Abstract

This paper analyses how cryptocurrency airdrops affect platform activity in decentralised finance (DeFi). Using daily active user (DAU) data from Arbitrum (March 2023) and Optimism (May 2022), I compare usage six months before and after each drop. Arbitrum's one-time event produced a sharp but temporary spike, while Optimism's staggered strategy sustained growth. Findings highlight how distribution design and governance incentives influence whether airdrops drive hype or long-term engagement.

1. Introduction

Decentralised Finance (DeFi) refers to using financial services like borrowing, sending, or lending money directly on a blockchain. It aims to remove traditional intermediaries such as banks and brokers, allowing users to have full control over their assets and reducing reliance on third parties. It uses smart contracts to verify every transaction, and transactions occur in a peer-to-peer manner. DeFi is growing to be increasingly popular and used globally. As of 2024, the total value locked (TVL) in DeFi protocols exceeds \$80 billion, and platforms handle billions of dollars in daily transaction volume. The number of unique DeFi users worldwide has surpassed 15 million, reflecting rapid adoption across regions. As a result, many projects compete for attention in this space; they often look for creative ways to attract users.

One popular method has been the use of airdrops, which is when crypto project developers distribute tokens with monetary value in order to encourage people to interact and bring publicity to their project. As Vitalik Buterin, co-founder of Ethereum, notes, "People respond strongly to free stuff, but free tokens alone don't guarantee long-term engagement." The two most common ways airdrops are distributed are either in a raffle style, where wallets are randomly selected, or where users must complete tasks before the airdrop, and the more tasks completed, typically means the more tokens are gained. Once users are granted their tokens, they can choose to hold them, trade them on exchanges, or use them within the project's ecosystem, depending on the token's functionality. However, in many projects, users often sell the tokens shortly after receiving them for quick profit. As a result, airdrops often generate only short-term hype, but studies show that this can often fail to translate into long-term user activity, consistent revenue generation, or meaningful growth in the protocol's economic ecosystem.

Knowing what truly drives user retention can help projects improve their strategic effectiveness. It allows them to avoid spending large amounts of capital on unsuccessful airdrops. While this study does not directly track individual user retention, it does measure whether overall activity levels increase or decrease after an airdrop. Understanding these usage trends allows projects to avoid spending large amounts of capital on unsuccessful campaigns. Sustained activity also boosts the stability of token economies because it prevents large, sudden sell-offs. This stability could help DeFi grow overall, as it may start to appeal to people looking for safer investment opportunities. The importance of studying user retention becomes clear when considering the risks protocols face if they fail to keep their communities active. Protocols may lack crucial liquidity to keep trading operations smooth if there's a lack of users. In severe

cases, some projects are forced to shut down entirely due to insufficient funding and low engagement. Additionally, poor user retention and large token sell-offs can damage a protocol's reputation and harm the credibility of its developers.

Given these risks, it's important to understand why some airdrops succeed in keeping users, while others do not. Successful airdrops manage to attract a loyal and active user base on their protocol, which maintains price stability and community engagement. In contrast, unsuccessful airdrops typically result in a short-term price spike, quickly followed by a large sell-off that causes the price to crash and leaves very little activity within the protocol, resulting in a flat price line.

Successful airdrops usually have methods to incentivise users to prolong their activity, such as distributing tokens in fixed-stage unlocks to prevent dumping. Recognising what separates successful airdrops from failed ones is key to both advancing academic knowledge and helping projects make better strategic decisions. Insights and data can help guide future protocols to build better-designed airdrops that encourage long-term retention and prevent token crashes. Additionally, most studies look at price changes; a focus on user retention may fill in a gap in blockchain data research.

In this research, I will analyse the effect of cryptocurrency airdrops on user retention in DeFi ecosystems. I will compare two notable airdrops in recent times: the Arbitrum airdrop and the Optimism airdrop. To conduct this analysis, I will extract on-chain statistical data, such as daily active users, using public APIs. By studying this on-chain data, I can directly measure whether users remained active after receiving airdrops, which will help reveal what truly drives long-term engagement in DeFi.

2 Literature Review

2.1 – Why do airdrops exist?

Airdrops are distributed to give recipients a financial stake in the ecosystem. They are considered a strategic tool designed to encourage recipients to become active, long-term users and draw them away from competing protocols. According to Xie et al. (2025), causal analysis shows that targeted distributions based on user behaviour correlate with higher sustained engagement.

2.2 - Short-Term Engagement vs. Long-Term Retention

Airdrops can drive rapid user acquisition by creating viral attention and incentivising wallet holders to explore a platform through monetary value. They are also relatively low-cost for developers. Xue et al. (2025) note that the Optimism airdrop didn't just bring in new users, but also sparked activity across different DAOs by connecting users through shared governance, helping Optimism grow within already active crypto communities.

2.3 - Airdrops and Platform Growth

While growth refers to the immediate spike in user activity following an airdrop, retention is about keeping those users engaged over time. Many projects succeed in generating short-term buzz but struggle to maintain momentum. Research suggests that how tokens are

distributed plays a key role in this outcome. Airdrops based on user behaviour—such as active participation or long-term involvement—and airdrops released in multiple rounds tend to perform better at sustaining engagement and limiting large sell-offs (Xie et al., 2025). Beyond token mechanics, the structure of the community also matters. Xue et al. (2025) highlight that protocols with strong DAO connections and built-in governance incentives are more likely to hold users' interest over time, turning early growth into lasting participation.

2.4 - What Drives Successful Airdrops?

Previous research has shown that airdrops have become a popular method to attract users, used by a growing number of developers. “Web3 projects have distributed billions of dollars of tokens at zero cost to users through airdrops” (Allen, 2023, p. 1). Similarly, Sie (2025) notes that airdrops help create an initial user base and spread tokens widely to jumpstart an ecosystem.

However, it is also well established that airdrops commonly lead to only a short-lived burst of attention and fail to drive real long-term user retention. “Many airdrops create sharp but temporary increases in wallet activity. Protocols report significant drop-offs in engagement within weeks of the airdrop” (Allen, 2023). Likewise, “a significant portion of recipients sell their tokens quickly, raising concerns about the effectiveness of airdrops in fostering genuine, lasting engagement” (Sie, 2025). These behaviours can contribute to price crashes, loss of credibility, and decreased liquidity—making it harder to attract investors and build a sustainable ecosystem.

Even though many airdrops fail to retain users, there are strong reasons they should drive platform growth. Airdrops attract publicity and, in theory, create a sense of loyalty—recipients may feel ownership. This can motivate users to stay active, participate in governance, and contribute liquidity, all of which support ecosystem health. As Allen (2023, p. 12) notes, “Airdrops are intended to convert recipients into ongoing users by giving them a financial stake in the network, encouraging participation and community loyalty.” Likewise, Sie (2025, p. 15) adds, “By distributing tokens broadly, projects hope to bootstrap networks and foster communities of users who remain engaged beyond the initial giveaway.” However, despite these intended outcomes, evidence shows that airdrops often fail to create lasting engagement—revealing a gap between theoretical goals and real-world results.

Given this mismatch, identifying the traits of successful airdrops becomes essential. Airdrops that reward users for specific tasks help filter recipients and attract users who are genuinely interested. Staged token unlocks discourage immediate selling and extend the impact of the airdrop. Sie (2025) notes that protocols using delayed token release saw reduced sell-offs and more protocol interaction. Some protocols even incentivise post-airdrop behaviour by offering bonus rewards or governance rights to users who remain active—helping transform one-time users into long-term participants.

3 Methodology

3.1 - Research Goal

This study investigates the effect of cryptocurrency airdrops on user retention in DeFi platforms by comparing Arbitrum and Optimism. This research aims to identify and understand what incentives and factors commonly contribute to a successful airdrop that retains users over an extended period of time. The centre of the methodology is extracting and studying chain data in order to evaluate user engagement over time to determine the success of both airdrops, Optimism and Arbitrum.

3.2 - Case Selection: Arbitrum and Optimism

Arbitrum and Optimism were selected because they each conducted two of the most high-profile airdrops in recent years. As both are established Layer 2 protocols within the Ethereum ecosystem, comparing the two ensures a fair and consistent analysis. Both projects share structural similarities in their airdrop design, such as tiered distribution and activity-based eligibility, which makes them highly comparable. Furthermore, their prominence means that there is a lot of on-chain data publicly available via API's, which can be extracted for thorough evaluation. Since both are in a similar time frame (2022-2023), external factors within the market are likely to be less of a limitation since they would have both experienced changes.

3.3 - Data Sources

The primary source of data for this research will be Dune Analytics - a blockchain data platform that allows custom queries to extract and isolate data. A critical measure that will be analysed is Daily Active Users (DAU). This shows the number of unique wallets interacting with the protocol each day. DAU was chosen as the core metric as it captures real user engagement, unlike price-based indicators, which can be affected by a mixture of factors.

3.4 - Data Collection and Processing

Data has been collected within a twelve-month time frame for both Arbitrum and Optimism, six months prior to the airdrop and six months post airdrop. This captures both the short-term volatility directly after airdrops and the longer-term aftermath of the airdrop post-volatility. The post-airdrop data can be compared directly to the data before the airdrop to thoroughly evaluate the effect of the airdrop on DAU. The airdrop dates used as reference points were March 23, 2023 (Arbitrum) and May 31, 2022 (Optimism). However, keep in mind that Optimism had 3 airdrops, and the reference date is the first one.

3.5 - Analytical Approach

Basic comparative statistics and graphing will be implemented once data is collected, such as logarithmic graphs, average DAU pre and post airdrop, percentage increases or decreases in DAU, and visual comparisons through a line graph with both airdrops plotted.

Percentage change will be calculated using the formula:

$$((\text{Post-airdrop DAU} - \text{Pre-airdrop DAU}) / \text{Pre-airdrop DAU}) \times 100$$
For example, if a protocol's average DAU increased from 80,000 to 160,000 after the airdrop, this would be a 100% increase.

3.6 – Sequence of methodology

1. Identify a consistent ± 180 -day window around each airdrop.

2. Pull DAU data through Dune using custom SQL queries.
3. Applied transformations (log scale and % change) to normalise trends.
4. Plotted user activity over time for each protocol.
5. Identified and compared post-airdrop retention patterns.
6. Paired user activity data with structural details of each airdrop, including token vesting schedules, eligibility criteria, and distribution mechanisms. This method does not track individual wallet retention but measures retention at the platform level using changes in DAU.

4 Analysis and Results

To better understand what drives successful airdrops, this paper analyses two of the most prominent examples in recent DeFi history: the Arbitrum and Optimism airdrops. Both distributed hundreds of millions of dollars worth of tokens to users and aimed to bootstrap long-term ecosystems through wide token distribution. Initially, both airdrops generated similar levels of growth in interest; however, their amount of daily active users took different trajectories over the longer term and the next 6 months post airdrops

4.1 - Arbitrum Airdrop Outcomes

After Arbitrum's airdrop on March 23 2023, the protocol saw a sharp spike in user activity, with DAU jumping from an average of 80,000 to a peak of over 600,000. While this surge shows strong initial interest, DAU activity quickly decreased and eventually settled at a baseline of around 160,000. Over the course of the 6 months and post spike, momentum gradually slowed and caused a downward sloping trend in DAU. Optimism, on the other hand, used a smaller token allocation for each individual drop but distributed it across three separate airdrops over time. This repeated distribution helped the protocol maintain user interest and drive recurring waves of activity. Rather than relying on a single moment of hype like Arbitrum, Optimism's staggered approach allowed it to continually re-engage users and sustain higher levels of daily activity across a longer period, which led to longer and sustainable growth.

The Arbitrum airdrop distributed 1.275 billion ARB tokens (12.75% of the total supply) to roughly 625,000 eligible wallets. To qualify, users had to perform on-chain activities like bridging tokens or participating in governance. Token allocations were tiered, meaning that more engaged users received higher amounts. While this design aimed to reward loyal participants and promote decentralisation, it did not fully prevent a post-airdrop drop-off. Despite the initial growth in DAU, Arbitrum struggled to maintain that momentum as usage declined in the weeks after the airdrop, and the platform failed to stabilise at the new peak. This suggests that although the drop was well-designed on paper, it may not have been sufficient to secure long-term retention without increasing community incentives on top of the 12.75% distributed.

4.2 - Optimism Airdrop Outcomes

Contrastingly, Optimism's airdrop strategy had a more sustained effect. While the first airdrop was launched on May 31st, 2022, it took a few days to show impact, with a major spike in DAU occurring on June 5th and rising from an average of under 10,000 to a peak of over 100,000 DAU. Although activity dipped shortly after, Optimism followed up with two additional

airdrops in the months that followed, where each brought new spikes in engagement. These recurring events helped maintain attention and steadily grow their user base. By the end of the six months, DAU had climbed to over 30,000, which is more than triple its baseline before the first airdrop. Rather than fading out, Optimism’s airdrops gradually built lasting growth, which may explain why it ended up retaining users more effectively than Arbitrum.

Optimism’s airdrop distributed 200 million OP tokens, 5% of the total supply, to around 250,000 eligible wallets. To qualify, users had to show early and consistent on-chain activity. A tiered bonus system was used to reward deeper engagement, with categories like “Gas Guzzler,” “Active Optimist,” and “Active Delegator.” The more categories users qualified for, the higher their token allocation was. Unlike Arbitrum, Optimism’s airdrop had no vesting schedule, and tokens were fully claimable from day one. Optimism's multi-staged airdrop structure helped reduce mass token dumping, which commonly happens from a lack of a vesting schedule. The fact that Optimism used a much smaller percentage of their total supply in comparison to Arbitrum emphasises Optimism's efficiency within their airdrops.

4.3 - Comparative Results

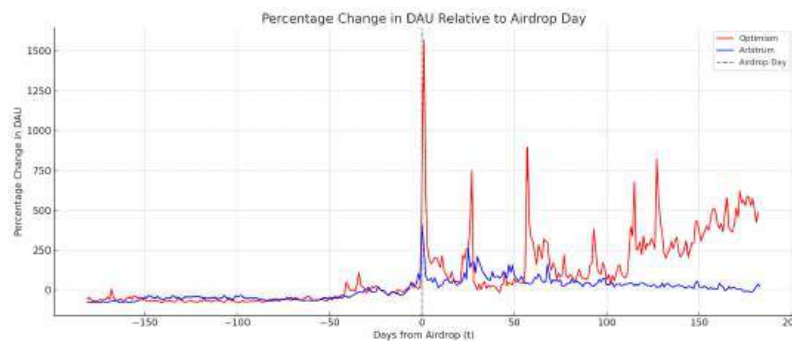


Figure 1: Percentage change in daily active users (DAU) relative to airdrop day for Optimism and Arbitrum.

Note: The Y-axis shows the percentage change in DAU compared to pre-airdrop baselines, while the X-axis represents time in days relative to the airdrop date (Day 0 = airdrop day). (% change calculated against the average DAU in the 30 days before the airdrop.)

4.4 - Airdrop Design Details

Although both projects used similar tiered criteria to distribute tokens to retain loyal users, a difference was that Arbitrum used a lengthy 4-year vesting schedule, allowing only a small portion of tokens to be released each month until March 16 2027. Contrastingly, Optimism’s airdrops demonstrated stronger retention and long-term engagement. While the first OP drop on May 31, 2022, triggered an initial spike to over 100,000 DAU, it was followed by two more airdrops over the next year, where each caused renewed spikes in activity and helped maintain interest in Optimism. Optimism’s multi-stage strategy helped sustain attention over time. As a result, Optimism’s DAU stabilised at a higher post-airdrop baseline, rising from under 10,000 to consistently over 20,000 by the end of 2022. In terms of short-term retention, Arbitrum saw 86% of recipients remain active after one month, compared to Optimism’s 69%. However, Optimism outperformed Arbitrum in longer-term engagement, with 8th-week retention averaging

30% higher than Arbitrum's 27%. This suggests that Optimism's staggered distribution and repeated incentives may have led to more frequent participation across its ecosystem.

4.4.1 - Arbitrum's Tiered Distribution

Points	Tokens Received (ARB)
3	1,250 ARB
4	1,750 ARB
5	2,250 ARB
6	3,250 ARB
7	3,750 ARB
8	4,250 ARB
9	6,250 ARB
10	6,750 ARB
11	7,250 ARB
12+	10,250 ARB

Figure 2: Table of tokens awarded per point on Arbitrums point-based system.

Arbitrum's tiered distribution worked on a point-to-token-based system; the more points you collected, the higher your allocation grew. Points were earned through various criteria, and a minimum of 3 points was needed to become eligible for the airdrop. Points gradually became harder to obtain, yet granted larger rewards.

Category	Requirement	Points earned
Transactions (Max 3 Points)	Performed ≥ 4 transactions	+1
	Performed ≥ 10 transactions	+1
	Performed ≥ 100 transactions	+1
Transaction Volume (Max 3)	Transacted \$10k+	+1
	Transacted \$50k+	+1
	Transacted \$250k+	+1
Smart Contract Interactions	Interacted with 4+ contracts	+1
Multi-Week Usage	Active over 2+ weeks	+1
Multi-Month Usage	Active over 2+ Months	+1

Bridging Funds on Arbitrum One (Max 3)	Bridged \$10k+ Bridged \$50k+ Bridged \$250k+	+1 +1 +1
Bridged Funds on Arbitrum Nova	Bridged any amount to Nova	+1
Nova Transactions	3+ transactions on Arbitrum Nova	+1
Early Usage	Before August 2022	Any points collected are worth double
DAO Governance	Participated in governance voting	+1
Multi-chain Usage	Active on other L2s like Optimism, Polygon, zkSync	+1
Arbitrum Ecosystem Engagement	Interacted with certain protocols (e.g., GMX, Treasure)	+1

Figure 3: Point-based system used by Arbitrum to allocate tokens based on user activity and engagement.

While Arbitrum's point-based system effectively rewarded past user behaviour and brought in over 600,000 eligible wallets, it offered limited incentives for continued participation. Tokens were distributed based on past engagement before the airdrop, and there was no mechanism to sustain activity post-drop. In contrast, Optimism used a multi-phase airdrop strategy that spread rewards across three rounds, which encouraged ongoing and future involvement and as a result, reinforced long-term engagement within its community.

4.4.2 - Optimism's Three Airdrops

Optimism structured their distribution around certain user categories across three separate airdrops. Each category targeted a specific activity, such as early adoption, governance participation, or gas spending and rewarded users based on their frequency of activity. Instead of a one-time drop, Optimism's multi-phase system gradually incentivised sustained engagement over time, which was a different approach to encouraging long-term activity.

To better understand how Optimism maintained engagement over time, each of the three airdrops can be analysed individually as they all had their own unique goals, eligibility rules, and outcomes.

Optimism Airdrop 1 – Early Adopters & Positive-Sum Contributors (May 31, 2022)

Optimism's first airdrop was designed to reward meaningful participation in the Ethereum and Optimism ecosystem. Eligibility was broken into six categories: **OP Early Users** (early users who bridged funds to Optimism), **Repeat Users** (users with consistent on-chain activity), **DAO Voters** (users that participated in governance), **Multisig Signers** (contributors to decentralised infrastructure), **Bitcoin Donors** (users who donated), and **Priced-Out Users** (active wallets on Ethereum with no Optimism activity, included to broaden access). Users qualifying for multiple categories were granted overlap bonuses and multiplied rewards, which served as a motivating mechanism for users to increase activity. Additionally, Optimism openly

promised future airdrops worth 19% of the supply. Through their first drop, Optimism positioned itself as a long-term ecosystem builder and helped itself earn early credibility and successfully kept users engaged beyond the initial hype and price spike.

Total OP Tokens Distributed	214 Million OP
Total OP Supply at date of Airdrop	4.29 Billion OP
Share of Supply	5%
Number of eligible wallets	248,699
Token Claim Deadline	September 2022

Figure 4: Key distribution stats from Optimism's first airdrop

Optimism Airdrop 2 – Governance and Gas Usage Incentives (Feb 9, 2023)

OP's second airdrop was engineered to not attract new users but retain existing users and gain loyalty by further incentivising them. The airdrop focused on two main key components - Governance participation and gas fee spending. Focusing on governance participation allowed OP to encourage users to actively use their tokens and reduce users who are just passive holders, which helped strengthen their community. Users who delegated their tokens to governance received 2x the bonus of rewards. Rewarding gas usage meant Optimism was incentivising depth of engagement, helping convert passive users into active participants within the ecosystem who generate volume, data, and feedback for future scaling. Through focusing on these two attributes, Optimism made it clear that they valued long-term retention rather than short-term hype. Only 0.27% of the total OP supply was distributed (around 11.7 million tokens), by focusing on users who were already engaged, Optimism reinforced its commitment to building a reliable and values-driven ecosystem. Only 19% of OP's total token supply was reserved for airdrops across all time. This signalled to the community that rewards would be measured and purposeful in lower quantities. This careful design avoided excessive sales pressure and also built trust among their user base.

In order to further encourage volume of activity per user, this airdrop had a reward multiplier system where users could qualify for up to 4 categories, and each category gained a multiplication in their rewards. The system was stackable, and users could reach a maximum multiplication of up to 4x. The categories were:

1. Governance Delegation

Users who had delegated their OP tokens for governance before the snapshot received a 2x bonus.

2. Multiple Address Delegation

Users who delegated governance across more than one address received an additional 2x bonus.

3. High Gas Usage on Optimism

Users who had spent significant gas on Optimism (≥ 0.5 ETH) earned a 2x bonus.

4. Repeat Governance Participation

Users who participated in multiple governance votes or proposals (not just one-time voting) got a 2x bonus

Total OP Distributed	11,742,277 OP
Per cent of total Supply	0.27%
Amount of Eligible Wallets	307,965
Delegation rewards	$\geq 2,000$ OP-Days $\rightarrow$ eligible to up to 5,000 OP
Gas Fee Reward	$\geq \$6.10$ spent $\rightarrow$ 80% rebate up to \$500 OP
Bonus Multipliers (maximum multiplication of 4x)	1.05 $\times$ to 2.00 $\times$ based on 4 behaviours
Snapshot Date	January 20th, 2023

Figure 5: Key statistics and eligibility criteria from Optimism's second airdrop

Optimism Airdrop 3 – Governance Deepening (September 18, 2023)

Their 3rd and final airdrop of the 6-month period zeroed in on governance and had a similar objective to the second drop; however, they generously rewarded users who had stayed active on their ecosystem for this long, as they distributed more OP, and there were a lower amount of Eligible addresses. This airdrop underscored Optimism's long-term focus on governance quality and built further trust with its users, leading to more overall user retention in the long term. Eligible users had to delegate tokens for at least seven days and accumulate 18,000 OP-Days (e.g., 100 OP delegated for 180 days). If funds were delegated to a voting delegate, only 9,000 OP-Days were required. This design helped reinforce a strong governance culture within Optimism's ecosystem, creating a base of active, voting participants who are likely to shape and stabilize the ecosystem long-term.

OP Distributed	19,411,313 OP
Per cent of total supply	0.45%
Eligible number of addresses	31,870
Delegation Threshold	$\geq 18,000$ OP-Days (or $\geq 9,000$ to Voting Delegate)
Minimum Delegation threshold	≥ 7 days
Reward Formula	$(0.67 / 365)$ OP per OP-Day, max 10,000 OP per wallet
Voting Delegate Bonus multiplier	2x OP
Distribution Strategy	Auto distributed, no claim needed
Snapshot Period	January 20 - July 20, 2023

Figure 6: Key statistics and eligibility criteria from Optimism's third airdrop

These airdrops highlight how protocols have moved beyond simple token giveaways and have now begun using distribution as a strategic tool to influence user behaviour. Each drop was tailored to guide engagement and increase activity within each ecosystem. This shift reflects a broader evolution in how DeFi projects approach community-building by focusing less on short-term hype and more on value-driven participation.

5 Public Reception

Blockchain data shows how users interacted with each protocol and provides statistical data; however, public reaction across social platforms like X (formerly Twitter) and Reddit provides additional insight into how the community perceived each airdrop's fairness, impact, and trustworthiness.

5.1 - Arbitrum Community Reaction

The announcement of Arbitrum's airdrop in March 2023 caused enormous anticipation across crypto communities. This is because after months of speculation, the long-awaited token was finally confirmed, and users rushed to check their eligibility and estimate their allocation. Social media lit up with reactions. Reddit user u/danyelsuave shared:

"I got 8k tokens in total. Sold some for BTC/ETH, but will also hold a good chunk,"
Showing the initial excitement among eligible claimants (Reddit, r/CryptoCurrency).

However, this wave of enthusiasm and hype also led to frustration and disappointment. On launch day, high traffic overwhelmed Arbitrum's claim site. Reddit user u/aaronismagic reported:

"Logged in right on time, but the site didn't load for 20 minutes. By the time it did, gas fees were insane,"

showing the performance issues many users faced. These roadblocks not only delayed claims but also spiked transaction costs and worsened the experience for early participants due to an unfair disadvantage.

The confusion was compounded by controversy over eligibility criteria. Users who had actively bridged funds and interacted with dApps on Arbitrum for months found themselves disqualified or receiving unreasonably small allocations. This led to widespread debate over the fairness of the distribution model. Reddit threads like "Why did I get nothing?" are filled with frustrated posts from users questioning Arbitrum's point system and accusing the protocol of favouring large wallets and bots. Several users speculated that "airdrop farmers" exploited the system, which caused allocations to be in favour of a few high-volume addresses.

Adding to the chaos, scammers took advantage of the hype. According to a report by Scam Sniffer, over 273 phishing websites impersonating Arbitrum went live during the claim

period, including fake claim pages, Twitter impersonators like @arbitrum_launch, and phishing DApps. Security firms like Redefine and CertiK posted real-time alerts, warning users about fake airdrop links and urging caution with wallet approvals (Cointelegraph, 2023). On Reddit, users like u/DubZChronic also spoke about these concerns:

“Scammers are hoping to capitalize on the complexity of crypto and users are excited for free money.” (Reddit, r/CryptoCurrency)

Arbitrum’s airdrop launch captured both the excitement and chaos that often come with major token releases. The buzz and anticipation were undeniable, but it quickly became clear that unclear rules, technical issues, and a lack of communication made things frustrating for many users. The mix of high hopes and messy execution left users feeling conflicted, making it harder for Arbitrum to gain loyal users.

5.2 - Optimism Community Reaction

On the other hand, Optimism’s first airdrop sparked a relatively positive reaction with many users praising both the eligibility design and the simplicity of claiming rewards. Recipients described it as fair and values-driven rather than a random giveaway. Reddit user u/toepin shared:

“I checked the Etherscan link and it is there – awesome!”

The drop’s inclusivity also stood out. Some noted that even indirect contributors were recognised, such as those who had donated to Gitcoin or participated in DAO votes. As one Redditor explained:

“Even if you’ve never used Optimism, you might still be eligible ... for doing things like donating in Gitcoin grants rounds ... or voting in DAO governance.”

This broader recognition of diverse contributions helped attract users heavily, which was key in their first airdrop and efficiently allocated the 5% supply.

However, the rollout wasn’t entirely smooth. High demand on launch day caused strain on the website, making it hard to log on and delays. The Optimism team later admitted:

“Wow, what a day. OP Drop #1 had a turbulent launch, which we finally stabilised after more than five hours of non-stop work.”

Price action also drew criticism. OP’s value dropped around 40% within hours, sparking debates over whether future rewards should exclude recipients who sold immediately. Unlike Arbitrum, OP tokens were unlocked immediately. One user complained:

“Why should Optimism Collective continue rewarding these kinds of mercenary actors who will dump their tokens at first sight?”

This split between loyal, long-term contributors and quick sellers really highlighted one of the biggest challenges with airdrops and finding a way to make them inclusive while still ensuring the biggest rewards go to those who genuinely stick around and add value over time.

6 Discussion

6.1 - Why Optimism Outperformed Arbitrum

Initially, Arbitrum generated more immediate attention and had a very sharp spike in DAU early on; however, this was short-lived, and there was quite a rapid decline. On the other hand, Optimism had a slower but steady increase that led to higher growth, which was retained for a longer period. This is due to a difference in strategy — instead of opting for a singular high-profile drop, Optimism's lengthier and staggered approach led to much higher loyalty and activity within their ecosystem. This approach helped OP avoid the steep post-airdrop drop-offs and sell-offs seen in Arbitrum and instead gain more consistent engagement over time, slowly but surely.

By focusing on rewarding governance participation and sustained gas usage, Optimism sent a clear message: users who contribute to the health of the protocol will be continuously recognised and rewarded. A crucial mechanism implemented by OP was doubling rewards for governance delegates. This not only boosted participation in decision-making but also reinforced the idea that Optimism's future would be shaped by its most committed members.

This, combined with a cautious allocation strategy of just 5.27% of the total supply spread across three drops compared to Arbitrum's 12.75% in one, signalled to the community that rewards would be measured, sustainable, and not prone to inflation. This helped Optimism build trust with loyal users and increase prolonged engagement. OP's approach appears to be more efficient as they allocated less capital but attracted more users. In the end, Optimism's layered, values-driven model proved more effective at keeping users engaged beyond the initial hype.

6.2 - What Makes an Airdrop Work

Prior evidence reveals that airdrops are most effective when they are explicitly designed to reward genuine contributors rather than random wallets in the ecosystem. The strongest results come from targeting users who are already engaged in the ecosystem and giving them clear, ongoing reasons to stay active — contrasting the common high-profile approach that lives off hype. This can be achieved by incentivising behaviours that strengthen the protocol, such as active governance participation, consistent transaction activity, and providing liquidity.

Tiered distribution models help separate and reward the most committed users, while bonus multipliers for specific behaviours like consistent governance delegation, high-volume transactions, or cross-protocol engagement encourage higher motivation and reward those who do well for the health of the ecosystem. Staggered approaches, either through multiple drops or vested unlock schedules, help sustain interest over time and prevent large-scale sell-offs immediately after distribution.

Allocating a fixed percentage of the total supply to airdrops over the long term also ensures that rewards remain valuable and strategically timed, rather than being wasted in a single event that only causes an attention spike. Combining clear eligibility requirements with meaningful rewards helps build trust, while openly communicating the purpose of the drop shows that the project is guided by long-term values. Additionally, when incentives are aligned with the ongoing health of the ecosystem and rewards feel genuinely earned, users are far more likely to stay engaged and continue contributing well after the initial excitement fades.

6.3 - Why Airdrops Fail

Airdrops appear to perform worse when they prioritise scale over strategy and give away large portions of the token supply in a single event without a plan to sustain momentum. Distributing too much at once can devalue the token and cause an immediate flood of sell-pressure, leaving little room for future incentives and users being left at the end of the sell-off and losing money. In these cases, the short-term hype is quickly replaced by a steep drop in engagement as opportunistic recipients take profits and exit. Another critical flaw is failing to differentiate between high-value contributors and passive recipients. When all wallets are treated equally, there is no added motivation for the most engaged community members to remain active, and potential long-term contributors are lumped in with passive users, which causes them to be demotivated and draws them away from the protocol.

Poor timing can also undermine an airdrop's potential, and launching during a bear market, a market crash, or before the protocol's core features and community are ready can cause new users to disengage quickly. Many failed airdrops also lack a structured follow-up plan to keep recipients involved. Additionally, without post-airdrop initiatives such as product updates, governance opportunities, bonus rewards, or additional campaigns, users have no reason to return after claiming their tokens.

6.4 - Broader Implications

Neglecting community desires during the planning stage can also be an issue; if recipients feel excluded, under-rewarded or misled, the protocol's credibility suffers. In some cases, poor eligibility checks allow bot networks or multi-account farmers to claim rewards, which diverts tokens away from genuine users who could have added long-term value. This not only wastes resources but can also create the perception that the project is careless or poorly managed. Ultimately, a lack of scarcity, poor targeting, and the absence of sustained engagement strategies can turn what should be a growth opportunity into a short-term distraction with lasting reputational damage.

Works Cited

- Allen, D. W. E. (2023, May 25). Crypto airdrops: An evolutionary approach (Working paper). Royal Melbourne Institute of Technology.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4440272
- Allen, D. W. E., Berg, C., & Lane, A. M. (2022). Why airdrop cryptocurrency tokens? Journal of Business Research (forthcoming). Royal Melbourne Institute of Technology.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4269239
- Arbitrum Foundation. (2023, March 16). Arbitrum airdrop eligibility and distribution.
<https://docs.arbitrum.foundation/airdrop-eligibility-distribution>
- BitDegree. (2023, March 21). Arbitrum airdrop: Eligibility, token allocation, and claim guide.
<https://www.bitdegree.org/crypto/tutorials/arbitrum-airdrop>
- BitDegree. (2023, June 5). Optimism airdrop: Eligibility and claims.
<https://www.bitdegree.org/crypto/tutorials/arbitrum-airdrop>
- Boxmining. (2023, February 10). Optimism OP token airdrop.
<https://boxmining.com/optimism-op-token-airdrop/>
- CoinDesk. (2023, June 7). What is Optimism?
<https://www.coindesk.com/learn/what-is-optimism>
- Cointelegraph. (2023, March 23). Beware of fake Arbitrum airdrops, community warns.
<https://cointelegraph.com/news/beware-of-fake-arbitrum-airdrops-community-warns>
- Cointelegraph. (2023, September 20). Optimism token dives 40% as proposal to bar dumpers from airdrops floated.
<https://cointelegraph.com/news/optimism-token-dives-40-as-proposal-to-bar-dumpers-from-airdrops-floated>
- Galaxy Research. (2023, July 14). Optimism vs. Arbitrum Part 1: Growth and ecosystem dynamics. <https://www.galaxy.com/insights/research/optimism-arbitrum-pt1-growth>
- Kaishi, N. (2023, March 8). Airdrops: A guide to token distribution & rewards. Medium.
<https://medium.com/@kaishinaw/airdrops-a-guide-to-token-distribution-rewards-7342271b328d>
- Nansen. (2023, March 20). An on-chain distribution model for the Arbitrum community.
<https://www.nansen.ai/research/an-on-chain-distribution-model-for-the-arbitrum-community>
- Optimism Foundation. (2022, May 31). Optimism token airdrop #1 overview.
<https://community.optimism.io/op-token/airdrops/airdrop-1>
- Optimism Foundation. (2023). OP token overview.
<https://community.optimism.io/op-token/op-token-overview>
- Reddit. (2023, March 23). Scam warning: We have seen dozens of Arbitrum scams today [Post]. r/CryptoCurrency.
https://www.reddit.com/r/CryptoCurrency/comments/11vvp9z/scam_warning_we_have_seen_dozens_of_arbitrum/
- Reddit. (2023, September 14). Optimism drop into my Coinbase Wallet app [Post]. r/ethereum.
https://www.reddit.com/r/ethereum/comments/16mox92/optimism_drop_into_my_coinbase_wallet_app/

- Sie, P. (2025). Regulating cryptocurrency airdrops. *Northwestern Journal of Technology and Intellectual Property*, 23. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4658712
- The Defiant. (2023, March 16). Arbitrum airdrop set for March 23. <https://thedefiant.io/news/defi/arbitrum-airdrop-set-for-march-23>
- Tokenomist. (2023, March 21). Arbitrum airdrop. <https://tokenomist.ai/arbitrum>
- X (formerly Twitter). (2023, March 23). Arbitrum official announcement [Tweet]. <https://x.com/arbitrum/status/1903915223486021697>
- Xangle Research. (2023, August 9). Analyzing the effectiveness of airdrops: Optimism vs. Arbitrum. <https://xangle.io/en/research/detail/1192>
- Xie, T., Ge, Y., & Kannan, K. (2025). Crypto airdrop success blueprint: A high-dimensional causal study using double machine learning. SSRN. <https://ssrn.com/abstract=5215263>
- Xue, L., Tian, J., An, J., & Zhao, X. (2025). Unlocking governance interlocks in blockchain DAOs: Insights from a token-airdrop context. SSRN. <https://ssrn.com/abstract=5262041>

CRISPR's Ability: The Different Systems Used to Edit DNA and RNA

By Arnav Jagtap

Abstract

Ever since its discovery Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) has been at the forefront of new and evolving medical technologies because of its ever growing potential in gene-editing and therapeutics. CRISPR works by using a protein such as Cas9 which is guided by a guideRNA to locate and cut parts of DNA. CRISPR can give scientists the ability to remove, replace, or modify genes which makes it one of the most transformative tools available to biologists. In this review, we will cover the differences and similarities between five major Cas proteins, detailing each protein's unique features and target specificities. Using a literature-based review, five different Cas proteins were compared to analyze their development, current applications, therapeutic potential, and their possible drawbacks. Each protein has different functions that allow them to recognize and modify specific genetic sequences, but each protein also presents potential drawbacks and unintended consequences on the genome. Understanding the capabilities and differences between these Cas proteins will be extremely important for the safety and effectiveness of CRISPR technologies.

Introduction

The discovery of CRISPR happened in 1987 when Yoshizumi Ishino and his colleagues at Osaka University noticed that there was an unusual repetitive DNA sequence in *Escherichia coli*. Later in 1995 Francisco Mojica observed similar sequences in archaea and believed that they were part of a prokaryotic immune system, which could defend those cells against viruses. Over the next 20 years scientists such as Emmanuelle Charpentier started to understand the capabilities of CRISPR including how Cas9, the CrisprRNA (crRNA), and the tracer RNA (tracrRNA) work in the CRISPR system. Later, Charpentier working with Jennifer Doudna, found out a way to combine the crRNA and the tracrRNA into a single guide RNA (sgRNA), which allowed for the base principles for practical genome editing. In 2013 multiple laboratories including ones led by Jennifer Doudna (Berkley) and Feng Zhang (Broad institute), were able to successfully use the CRISPR-cas9 system to edit human (*ex vivo*) and animal cells, showing the groundbreaking advances that CRISPR has made. These advances include disease treatment, viral detection, and even a potential for *in vivo* gene editing for humans. (Gostimskaya)

CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats, which was derived from the immune system that is found in bacteria and archaea. In bacteria and archaea, CRISPR stores fragments of viral DNA that it collects when the cell is attacked, so that when the same virus attacks one more time the spacer is able to guide a Cas protein like Cas9 to recognize the virus and cut up the invading DNA. When used in a laboratory, Cas9 can act as molecular scissors as it is being guided by the sgRNA to a DNA sequence and it is able to introduce a double strand break in the DNA. The cell can then repair the cut and it either changes, repairs or adds genes where that double strand break occurred. CRISPR is extremely important because it allows for control of genetic material which allows for the development of

therapies for genetic diseases like sickle cell anemia. CRISPR's efficiency compared allows for it to become a transformative technology in medicine and in biotechnology. (National Institute of General Medical Sciences)

Cas9 is one of the most famous and most widely used Cas protein as it is guided by a small RNA called the single guideRNA (sgRNA) to a specific DNA sequence where it makes a double stranded cut. This makes it extremely versatile for many different scenarios which makes it one of the most widely used Cas proteins. (Mali) Cas3, on the other hand, does not cut DNA at a single spot it binds to its targets and then it starts to degrade DNA sometimes for thousands of bases. Although this makes it extremely useful for removing large sections of DNA and for antiviral applications it makes it extremely impractical for it to be used for more precise cuts, as it has the potential to cause major damage to the genome. (Sinkunas) Cas12a makes staggered cuts which are called "sticky end" cuts instead of just making 2 blunt ends in the DNA like Cas9. Since Cas12a is extremely small and doesn't need a tracrRNA, it is much easier to deliver in smaller delivery vectors. (Broughton) Cas14 is one of the few Cas proteins that specifically just target single stranded DNA which makes it perfect to be able to detect rare genetic disorders with very minimal error, while also being extremely small. (Savage) While all of these Cas proteins are extremely useful for DNA, Cas13 is one of the few Cas proteins that are able to be used in RNA which makes it able to degrade viral RNA without having to change the genome. (Huang) Although there are more than 30 Cas proteins that have been discovered, these 5 proteins were focused on because of their unique skillsets to change genomic sequences like DNA, ssDNA and RNA (Table 1).

Cas Protein	Target	Size of Protein	Unique Features	Limitations	Current Uses
Cas9	Double Stranded DNA	1300 amino acids	Different variants that have different features	Off-target cleavage	Genome editing and therapy
Cas12a	Double Stranded DNA	1200 amino acids	Can make staggered cuts	Not as good for all targets	DNA editing and diagnostics
Cas13	Single Stranded RNA	1300 amino acids	Specifically cleaves RNA	Can only cleave RNA	RNA cleavage and antivirals
Cas14	Single Stranded DNA	550 amino acids	Specifically cuts single stranded DNA	Can only cleave ssDNA	DNA detection and editing
Cas3	Double Stranded DNA	1100 amino acids	Can cleave large stretches of DNA	Damages the whole genome with massive cuts	Large scale DNA deletions

Table 1: Comparison of different CRISPR Cas Proteins. This table shows the different targets, sizes, special abilities, limitations and uses of various CRISPR Cas proteins. Cas9 and Cas12a mostly focus on double stranded DNA where Cas9 cuts with blunt end and Cas12a is known for its staggered cuts. Cas13 and Cas14 give specific targeting for single stranded RNA and DNA, respectively. Cas3 is known for being able to make large scale genomic deletions. All of these Cas proteins give CRISPR the ability to be used in high precision gene therapy or rapid diagnostics.

Despite the recent discoveries about CRISPR there are significant gaps in our understanding of its capabilities, and the potential downsides of using CRISPR. These downsides include the off-target effects and the long term side effects of using CRISPR based therapies. Additionally there is already a lot of research that has been conducted on Cas9 but there is still mystery regarding different Cas proteins especially understanding their precision, and efficiency. The main goals of CRISPR research is to achieve precise and efficient genome editing that can develop gene therapies for many genetic disorders and to enhance genetic disease detection. By truly understanding the different unique features of those 5 Cas proteins is extremely important for reaching these goals, as they each offer mechanisms for degrading sequences or targeting DNA or RNA. If researchers choose the right tools then they can maximize efficiency and push the boundaries of what CRISPR based technologies can achieve.

Cas9

Mechanism

Cas9 is a Class 2 Type II RNA-guided endonuclease which binds a guide RNA (gRNA) that guides it to a complementary DNA sequence. Before cutting, the Cas9 protein recognizes a short DNA motif which is called the protospacer adjacent motif (PAM), which is needed for binding (Figure 1A). Once it is bound, it unwinds the DNA so the guide RNA can pair with its target which forms an R-loop which confirms that the gRNA has identified the correct target sequence while also positions the Cas9 enzymes nuclease domains (HNH and RuvC) to do the double stranded break. (Li) After the break, the cell's repair machinery, non-homologous end joining or homology directed repair, fixes the breaks which allows for deleting, insertions or precise edits. CRISPR-induced double stranded breaks are repaired by 2 different DNA repair mechanisms including Non-Homologous end Joining (NHEJ) and Homology-directed repair (HDR). NHEJ rejoins the double stranded breaks by joining the ends which can introduce small insertions or deletions. HDR uses homology from a donor template to copy the exact sequence into the break which can allow for gene correction with high accuracy (Figure 1B). (Asmamaw)

Limitations

The CRISPR/Cas9 faces many challenges that can limit its therapeutic application including delivery, off-target effects and immune stimulation. Starting with delivery, Cas9 and the gRNA have to be efficiently transported into the target cell, but the common methods that are used such as viral vectors are difficult to use because of the size constraints. There are issues with achieving cell-type specificity because imperfect targeting can lead to the editor reaching unintended tissues within the body. Adding on, there are also off-target effects, where Cas9 can cut DNA at unintended sites which can cause harmful mutations, this risk increases when there are mismatches between the guide RNA and the DNA. Lastly, immune stimulation is a concern because the Cas9 protein which is taken from bacteria could be recognized as foreign by the human immune system which may create safety risks. (Karimi)

Current Uses

CRISPR can be used in research to modify, or delete genes or create disease models in cells or animal systems which can help us understand genetic pathways and biological processes. CRISPR is also being used in clinical applications especially in therapies for sickle cell disease where patients' hematopoietic stem cells are edited *ex vivo* to increase the fetal hemoglobin production to reduce the disease symptoms. (FDA) CRISPR is also being used to improve crop traits, and optimize livestock genetics which shows the potential outside of human medicine. All of these current use cases of CRISPR lay the foundations of what CRISPR could accomplish as it has already bridged the gap between fundamental theories and actual application. (Ansori)

Future Therapeutic Applications

As mentioned before the potential of CRISPR/Cas9 gene editing is limitless but it is currently moving towards *in vivo* therapies that can permanently correct disease with high precision and accuracy. Researchers are currently developing advanced Cas9 variants which include PAM expanded versions to reduce the current off-target risks. One of the most important parts of the delivery are the vectors and researchers are working on nonviral vectors that can improve specificity which would be extremely important for effective treatments. There is potential for a future with one-time gene-editing treatments that can give durable outcomes for a lot of different diseases which are being worked on as scientists try to refine the editing strategies and delivery systems. For one-time gene-editing treatments to be possible there would have to be high on-target precision with Cas9 along with extremely efficient delivery vectors. This is considered possible because other Cas proteins have already shown that they can correct mutations and there is work being done to produce smaller and more precise nuclease variants. (Li)

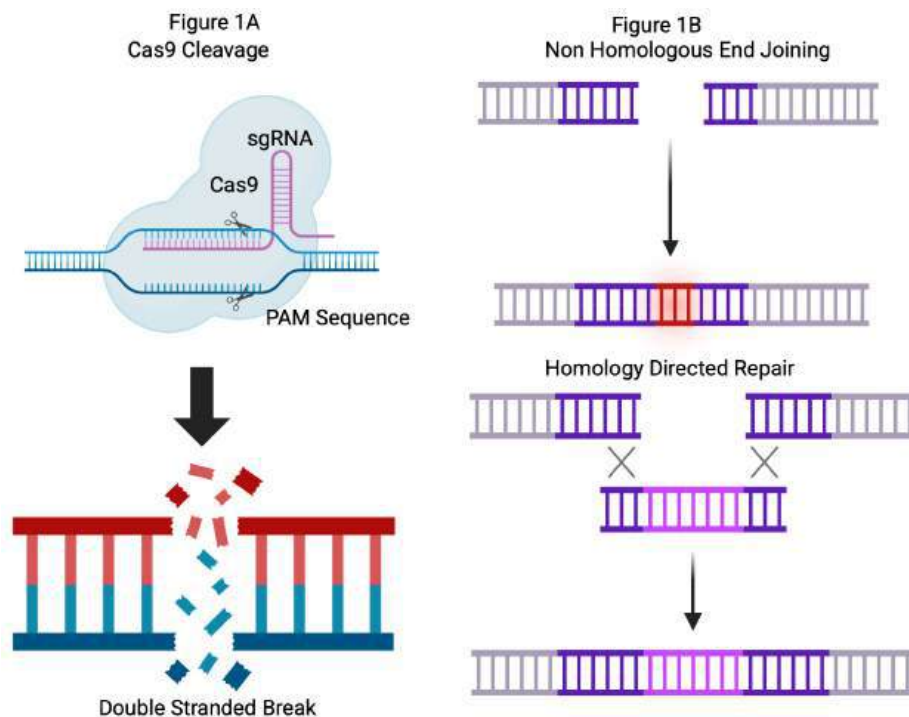


Figure 1: Mechanism of CRISPR-Cas9 Cleavage and Repair Pathways. **A)** Cas9 cleavage shows the targeted destruction of DNA. The DNA Cas9 is guided to the target gene sequence by the single gRNA. The Cas9 can confirm the binding by recognizing the adjacent PAM sequence before creating the Double Stranded Break. **B)** The illustration on the top shows Non-Homologous End Joining which joins the broken DNA ends. This will often result in small insertions or deletions which can disrupt the gene. The illustration on the bottom shows Homology-Directed Repair. This is a more precise repair pathway that uses an external Repair template which is shown in purple, which allows the cell to use the template to incorporate specific new genetic material for gene correction.

Cas12a

Mechanism

Cas12a is a Class 2 Type V RNA guided endonuclease that is used for genome editing. While Cas9 needs a sgRNA which includes a crRNA and a tracrRNA, Cas12a only needs a crRNA and a single RuvC nuclease domain that can make a staggered double stranded break, with the 5' end overhanging, while also being able to degrade single stranded DNA (ssDNA). Cas12a can recognize T-rich PAM sequences that are upstream of the target site, and it can unwind the DNA and form an R-loop with the crRNA to guide the cleavage (Figure 2). Cas12a's PAM sequence is optimized to form contacts with Thymine bases; the shape of T bases binds with Cas12a better than G/C bases. It can also process its own pre-crRNA into mature crRNAs without needing tracrRNA simplifying its function. This allows Cas12a to be extremely efficient for genome editing. (Paul)

Limitations

Just like Cas9, Cas12a has limitations that can affect its precision. Cas12a can cause off-target cleavage at DNA sequences that are similar to the intended target. It also has an indiscriminate ssDNA which means that after binding, Cas12a could cut nearby DNA, interfering with important cellular processes. (Marugan) The activity of Cas12a is dependent on the presence of a specific PAM which can limit the parts of the genome that it can target. Cas12a requires a specific T-rich PAM located upstream of the target sequence which is much more restrictive than Cas9's PAM and can limit the sites where Cas12a can target, especially in GC-rich regions.

Current Uses

Cas12a is currently being used for both genome editing and molecular diagnostics. In genome editing, Cas12a enables precise gene knockouts and targeted mutagenesis (making a permanent mutation in a specific part of the genome) in bacteria, plants, and mammalian cells. Because it is able to process multiple crRNA's, it is able to simultaneously edit many genes at the same time. (Paul) Cas12a is being used in diagnostics through the DETECTR (DNA Endonuclease Targeted CRISPR Trans Reporter) assay which uses the specificity of Cas12a. Cas12a is programmed with a guide RNA to bind to a specific DNA sequence from a pathogen. Once the target is recognized, Cas12a is activated and it exhibits its collateral cleavage of all nearby ssDNA. (Gostimskaya) This cleavage cuts a fluorescent ssDNA reporter, which produces a detectable signal which shows the presence of the target DNA. This is what allowed Cas12a to be extremely useful in COVID-19 detection. (Rananaware)

Future Therapeutic Applications

Researchers are working on improving Cas12a for future therapeutic applications. There is work being done to produce variants that have more relaxed PAM requirements which can allow more of the genome to be targeted. These different variants could be used to correct disease causing mutations or even to treat genetic disorders. Controlled off-target and the ssDNA activity would make Cas12a much safer for *in vivo* applications. (Ma)

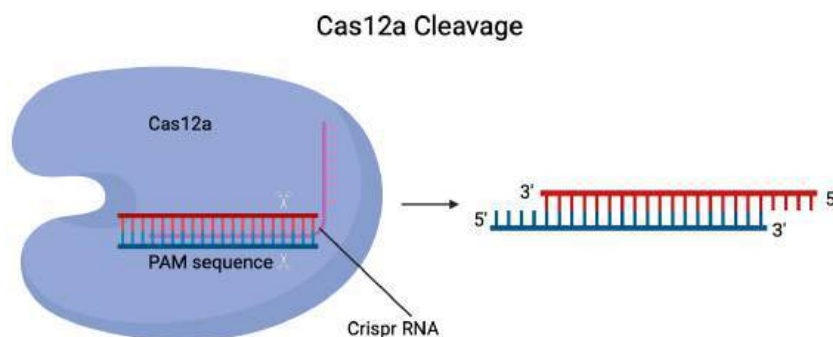


Figure 2: Mechanism of Cas12a Cleavage. The Cas12a system uses its T-rich PAM sequence on the non-target DNA strand so that it can anchor the complex and start the R-loop formation. This allows the Cas12a enzyme to start the staggered cut which results in the 5' overhang that results from Cas12a cleavage.

Cas13

Mechanism

Cas13 is a Class 2, Type VI CRISPR system that targets RNA rather than DNA. It uses a single crRNA to recognize complementary ssRNA targets, without the tracrRNA that is needed for Cas9. Once the crRNA has bound to the target, Cas13's HEPN (Higher Eukaryotes and Prokaryotes Nucleotide Binding) nuclease domains are activated which cleaves the RNA (Figure 3). Cas13 orthologs also have collateral RNase activity which cleaves other non-target RNA after activation. This can allow Cas13 to degrade target RNAs and create a detectable signal for diagnostics, such as in SHERLOCK (Specific High Sensitivity Enzymatic Reporter Unlocking). (Larsen)

Limitations

One of the major limitations of Cas13 is that it has collateral RNA cleavage which can degrade other non-target RNAs in the cell and can disrupt normal cellular processes. This non-specific activity can complicate editing applications. Also some Cas13 variants need a specific protospacer flanking sequence or show different levels of efficiency depending on the RNA structure. This can reduce the target selection and reduce the consistency of Cas13 treatments. (Allemailem)

Current Uses

Cas13 is used for quick RNA detection which includes pathogen and viral RNA diagnostics by using its target activated collateral cleavage to produce a measurable signal. SHERLOCK (Specific High-Sensitivity Enzymatic unLOCKing) is a diagnostic method that uses Cas13 which is guided by RNA to the target RNA. When it binds, Cas13 is activated and collaterally cleaves reporter RNA which has a fluorophore and a quencher, so when cleaved they separate and produce a fluorescent signal that can be easily detected. (Kellner) It can also be used for RNA knockdown and transcript regulation in mammalian cells which can be an alternative to RNA interference. Catalytically inactive Cas13 can allow for RNA binding without cleavage which can enable live-cell RNA imaging and tracking. These applications have expanded Cas13's role in research and other diagnostic platforms especially for detecting viral pathogens. (Savage)

Future Therapeutic Applications

Cas13 is considered the future of RNA targeted therapeutics as it is being explored for degrading disease associated RNAs such as viral genomes. RNA-level targeting with Cas13 is reversible which can reduce the risks that are associated with permanent DNA edits. There are a lot of efforts to improve specificity and reduce collateral damage, which includes optimizing the guide design and engineering different Cas13 variants. Cas13 can also evolve into a clinical diagnostic tool that can quickly detect disease-related RNAs in patient samples. (Zhu)

Cas13 Cleavage

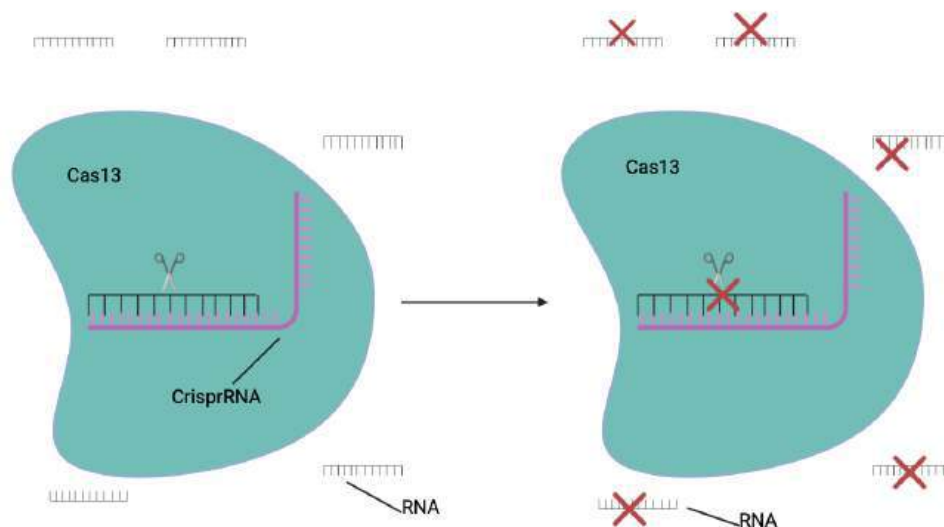


Figure 3: Mechanism of Cas13 RNA Collateral Cleavage. This figure shows the unique collateral cleavage of the Cas13 protein. After recognizing the target sequence using the crRNA the Cas13 enzyme binds to its target and starts cutting at that sight. This leads to a change that causes the protein to start cleaving every nearby RNA strand.

Cas14

Mechanism

Cas14 is a compact RNA guided nuclease that only targets ssDNA using a crRNA to recognize complementary sequences. After binding to its target, Cas14 changes its shape

activating the RuvC nuclease domain to cut the ssDNA. (NIGMS) Cas14 also exhibits collateral cleavage so it can cut nearby ssDNA molecules, allowing it to be highly specific while also generating detectable signals. (Harrington) This gives researchers the ability to know when Cas14 has successfully recognized the target DNA (Figure 4).

Current Uses

Cas14 is mainly used in molecular diagnostics because it has high specificity with ssDNA sequences. It is the main part of the Cas14-DETECTR assays which can detect pathogens and mutations in DNA samples. The collateral cleavage of Cas14 is paired with fluorescent reporter molecules to generate measurable signals. Its small size makes it suitable for biosensing platforms and point of care diagnostic tests. (Savage)

Limitations

Cas14 is limited because it can only target single stranded DNA which makes it ineffective for editing double stranded genomes. Its efficiency in mammalian cells is much lower than the efficiency of Cas9 or Cas12a which restricts its applications. Adding on, delivering Cas14 *in vivo* is challenging even though it has a small size. Because it requires a guide RNA that has to be correctly folded, and delivered to the same cells as Cas14 and because of the instability of RNA co-delivery can be inefficient. Lastly, because of its collateral damage it can induce off-target effects and the lack of testing in therapeutics can make it difficult to use clinically. (Ullah)

Future Therapeutics

The future of Cas14 is mostly in diagnostics and biosensing where it is able to use its ultra small size, PAM independent targeting and single stranded DNA cleavage. The DNA cleavage can be used for rapid and portable detection platforms. Research is currently ongoing to improve the specificity of Cas14 and working on point-of care devices that can detect pathogens, detect variants and recognize cancer biomarkers. There is research looking into Cas14 variants with expanded targeting capabilities which could broaden its potential for controlled genome cleavage. (Giquel)

Cas14 Cleavage

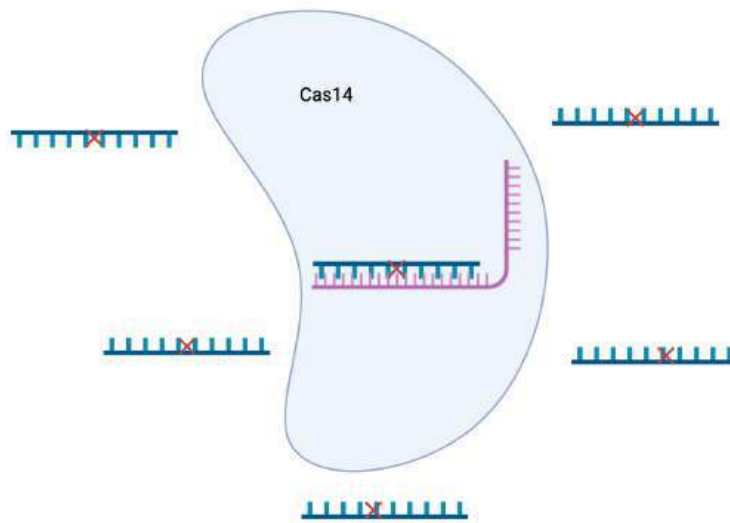


Figure 4: Mechanism of Cas14 Cleavage. This figure shows how the Cas14 protein can cleave single stranded DNA. Cas14 is a much smaller protein at roughly 550 amino acids and it has ability to perform cleavage of ssDNA without needing a PAM sequence. This shows how the sgRNA can bind to the target strand leading to collateral damage giving Cas14 a unique ability to be effective for molecular diagnostics and DNA detection.

Cas3

Mechanism

Cas3 is a multi-functional enzyme in the Type 1 CRISPR-Cas system that functions as a single stranded DNA nuclease and an ATP dependent helicase. This allows it to unwind and degrade DNA after target recognition by using the Cascade complex. Cas3 is recruited by the Cascade complex and undergoes conformational changes that activate Cas3's helicase and nuclease domains which allows for processive DNA degradation (Figure 5). This shows how Cas3 can put together target recognition and efficient DNA degradation to carry out immune defense. (Huo)

Current Uses

Cas3 has been used for genome engineering and targeted DNA degradation in bacterial and eukaryotic systems. It is separated from Cas9 because it is able to generate large and processive deletions which makes it extremely useful for removing large genomic regions, or conducting functional genomic screens. (Csorgo) Current applications currently include bacterial immunity studies, programmable DNA excision and diagnostic platforms that use Cas3 mediated DNA cleavage. (Hou)

Limitations

Cas3 has several limitations because of the major deletions that Cas3 makes. Since the deletions endpoints are difficult to control, the precision is extremely reduced compared to Cas9. Also editing efficiency is extremely variable as the multi-protein Cascade-Cas3 complex requires

a coordinated delivery which can be a challenge in eukaryotic cells. Lastly, the large protein requirements limit the therapeutic applications. (Yang)

Future Therapeutics

Cas3 processive DNA degradation gives potential for therapeutic applications, like correcting genetic disorders that require large deletions or removing pathogenic DNA. Research is working towards improving the delivery methods, controlling deletion boundaries and making it into a safe and programmable genome editing platform. This could make Cas3 a powerful method for gene recovery once the technical challenges are addressed. (Stroik)

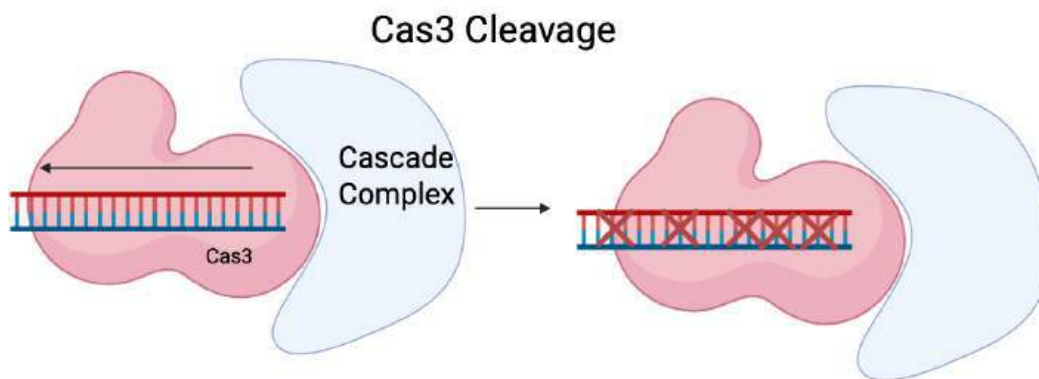


Figure 5: Mechanism of Cas3 Cleavage. Cas3 moves along the DNA helix like a Pac-Man after it is recruited by the Cascade complex. This results in the continuous shredding of the DNA strand which neutralizes the target.

Conclusion and Potential Outcomes

CRISPR Cas systems offer a very diverse and programmable set of options that originated from the immune system of bacteria. Although Cas9 is the most popular and most reliable Cas proteins for creating double stranded breaks in DNA, there are many other proteins such as Cas12a which can offer a staggered cut. Adding on there is Cas13 which can give researchers the ability to target RNA, while Cas14 can target ssDNA, on the other hand Cas3 has the ability to make large deletions. This gives scientists the choice of the specific pair of molecular “scissors” which are adapted for a target strand of specific size, or nucleic acid type.

There is also major potential for CRISPR proteins to transform precision medicine and biotechnology to save countless lives. By correcting genetic mutations at their source, CRISPR has already helped develop treatments for diseases that were thought to be incurable. These include sickle cell disease and hereditary blindness. Its ability to create rapid diagnostics for infectious disease and helping to engineer climate resilient crops makes it one of the most important scientific advancements of the 21th century.

Acknowledgements

This project was completed under the guidance of a PhD student mentor, Theresa Chen. All figures were created with BioRender.

Works Cited

- Gostimskaya, Irina. "CRISPR–Cas9: A History of Its Discovery and Ethical Considerations of Its Use in Genome Editing." *Biochemistry (Moscow)*, vol. 87, no. 8, Aug. 2022, pp. 777–88, <https://doi.org/10.1134/s0006297922080090>.
- "What Is CRISPR? | National Institute of General Medical Sciences." Nih.gov, 2024, nigms.nih.gov/biobeat/2024/10/what-is-crispr.
- Mali, Prashant, et al. "Cas9 as a Versatile Tool for Engineering Biology." *Nature Methods*, vol. 10, no. 10, Oct. 2013, pp. 957–63, <https://doi.org/10.1038/nmeth.2649>. Accessed 25 Apr. 2019.
- Sinkunas, Tomas, et al. "Cas3 Is a Single-Stranded DNA Nuclease and ATP-Dependent Helicase in the CRISPR/Cas Immune System." *The EMBO Journal*, vol. 30, no. 7, Feb. 2011, pp. 1335–42, <https://doi.org/10.1038/emboj.2011.41>. Accessed 5 Feb. 2021.
- Broughton, James P., et al. "CRISPR–Cas12-Based Detection of SARS-CoV-2." *Nature Biotechnology*, vol. 38, Apr. 2020, <https://doi.org/10.1038/s41587-020-0513-4>.
- Savage, David F. "Cas14: Big Advances from Small CRISPR Proteins." *Biochemistry*, vol. 58, no. 8, Feb. 2019, pp. 1024–25, <https://doi.org/10.1021/acs.biochem.9b00035>. Accessed 16 Oct. 2021.
- Huang, Zhanchao, et al. "CRISPR-Cas13: A New Technology for the Rapid Detection of Pathogenic Microorganisms." *Frontiers in Microbiology*, vol. 13, Oct. 2022, p. 1011399, <https://doi.org/10.3389/fmicb.2022.1011399>.
- Li, Tianxiang, et al. "CRISPR/Cas9 Therapeutics: Progress and Prospects." *Signal Transduction and Targeted Therapy*, vol. 8, no. 1, Jan. 2023, pp. 1–23, <https://doi.org/10.1038/s41392-023-01309-7>.
- Asmamaw, Misganaw, and Belay Zawdie. "Mechanism and Applications of CRISPR/Cas-9-Mediated Genome Editing." *Biologics : Targets & Therapy*, vol. 15, no. 1, Aug. 2021, pp. 353–61, <https://doi.org/10.2147/BTT.S326422>.
- Karimi, Mohammad Ali, et al. "Challenges and Opportunities in the Application of CRISPR-Cas9: A Review on Genomic Editing and Therapeutic Potentials." *Medical Principles and Practice*, Karger Publishers, July 2025, pp. 1–30, <https://doi.org/10.1159/000547334>.
- Food and Drug Administration. "FDA Approves First Gene Therapies to Treat Patients with Sickle Cell Disease." FDA, 8 Dec. 2023, www.fda.gov/news-events/press-announcements/fda-approves-first-gene-therapies-treat-patients-sickle-cell-disease.
- Ansori, Arif NM, et al. "Application of CRISPR-Cas9 Genome Editing Technology in Various Fields: A Review." *Narra J*, vol. 3, no. 2, Aug. 2023, pp. e184–84, <https://doi.org/10.52225/narra.v3i2.184>.
- Li, Tianxiang, et al. "CRISPR/Cas9 Therapeutics: Progress and Prospects." *Signal Transduction and Targeted Therapy*, vol. 8, no. 1, Jan. 2023, pp. 1–23, <https://doi.org/10.1038/s41392-023-01309-7>.

- Paul, Bijoya, and Guillermo Montoya. "CRISPR-Cas12a: Functional Overview and Applications." *Biomedical Journal*, vol. 43, no. 1, Feb. 2020, pp. 8–17, <https://doi.org/10.1016/j.bj.2019.10.005>.
- Murugan, Karthik, et al. "CRISPR-Cas12a Has Widespread Off-Target and DsDNA-Nicking Effects." *Journal of Biological Chemistry*, vol. 295, no. 17, Apr. 2020, pp. 5538–53, <https://doi.org/10.1074/jbc.RA120.012933>.
- Rananaware, Santosh R., et al. "Programmable RNA Detection with CRISPR-Cas12a." *Nature Communications*, vol. 14, no. 1, Sept. 2023, p. 5409, <https://doi.org/10.1038/s41467-023-41006-1>.
- Ma, Enbo, et al. "Directed Evolution Expands CRISPR–Cas12a Genome-Editing Capacity." *Nucleic Acids Research*, vol. 53, no. 13, Oxford University Press (OUP), July 2025, <https://doi.org/10.1093/nar/gkaf649>. Accessed 20 Sept. 2025.
- Larsen, Benjamin B., et al. "RNA Structure Modulates Cas13 Activity and Enables Mismatch Detection." *Nature Biotechnology*, *Nature Portfolio*, Oct. 2025, <https://doi.org/10.1038/s41587-025-02868-6>.
- Allemailem, Khaled, et al. "Current Updates on the Structural and Functional Aspects of the CRISPR/Cas13 System for RNA Targeting and Editing: A Next-Generation Tool for Cancer Management (Review)." *International Journal of Oncology*, vol. 66, no. 5, Spandidos Publications, May 2025, pp. 1–21, <https://doi.org/10.3892/ijo.2025.5748>.
- Kellner, Max J., et al. "SHERLOCK: Nucleic Acid Detection with CRISPR Nucleases." *Nature Protocols*, vol. 14, no. 10, Sept. 2019, pp. 2986–3012, <https://doi.org/10.1038/s41596-019-0210-2>.
- Zhu, Guanglin, et al. "CRISPR-Cas13: Pioneering RNA Editing for Nucleic Acid Therapeutics." *BioDesign Research*, vol. 6, American Association for the Advancement of Science, Jan. 2024, <https://doi.org/10.34133/bdr.0041>.
- Harrington, Lucas B., et al. "Programmed DNA Destruction by Miniature CRISPR-Cas14 Enzymes." *Science*, vol. 362, no. 6416, Oct. 2018, pp. 839–42, <https://doi.org/10.1126/science.aav4294>. Accessed 15 Mar. 2021.
- Savage, David F. "Cas14: Big Advances from Small CRISPR Proteins." *Biochemistry*, vol. 58, no. 8, Feb. 2019, pp. 1024–25, <https://doi.org/10.1021/acs.biochem.9b00035>.
- Ullah, Numan, et al. "Comparative Analysis and Phylogenetic Insights of Cas14-Homology Proteins in Bacteria and Archaea." *Genes*, vol. 14, no. 10, Multidisciplinary Digital Publishing Institute, Oct. 2023, pp. 1911–11, <https://doi.org/10.3390/genes14101911>. Accessed 11 Oct. 2024.
- Benoit Giquel. "CRISPR-Cas14: A Family of Small DNA-Targeting Enzymes Enabling High-Fidelity SNP Genotyping." *Addgene.org*, 2018, blog.addgene.org/crispr-cas14-a-family-of-small-dna-targeting-enzymes-enabling-high-fidelity-snp-genotyping.
- Huo, Yanwu, et al. "Structures of CRISPR Cas3 Offer Mechanistic Insights into Cascade-Activated DNA Unwinding and Degradation." *Nature Structural & Molecular Biology*, vol. 21, no. 9, Aug. 2014, pp. 771–77, <https://doi.org/10.1038/nsmb.2875>.

- Csörgő, Bálint, et al. “A Compact Cascade–Cas3 System for Targeted Genome Engineering.” *Nature Methods*, Oct. 2020, pp. 1–8, <https://doi.org/10.1038/s41592-020-00980-w>.
- Hou, Zhonggang, et al. “Introducing Large Genomic Deletions in Human Pluripotent Stem Cells Using CRISPR-Cas3.” *Current Protocols*, vol. 2, no. 2, Feb. 2022, p. e361, <https://doi.org/10.1002/cpz1.361>. Accessed 21 Feb. 2022.
- Yang, Peihong, et al. “Research Progress on the Mechanism and Application of the Type I CRISPR-Cas System.” *International Journal of Molecular Sciences*, vol. 25, no. 23, Multidisciplinary Digital Publishing Institute, Nov. 2024, pp. 12544–44, <https://doi.org/10.3390/ijms252312544>. Accessed 19 Jan. 2025.
- Stroik, Susanna. “CRISPR 101: Cas9 vs. the Other Cas(S).” *Blog.addgene.org*, 8 Dec. 2022, blog.addgene.org/crispr-101-cas9-vs.-the-other-cas.

Urban Heat Vulnerabilities in Mid-Sized Cities: A Case Study of Lansing, Michigan

By Haneesha Padavala

1. Abstract

Urban heat islands (UHIs) are areas of land that are significantly warmer than their surrounding areas. People living in UHIs are more prone to poor air quality, heat-related deaths, and higher energy costs. UHIs can also contribute to ecological strain and can unevenly affect vulnerable populations. For this study, Lansing, Michigan, is used as a case study due to its characteristics as a mid-sized city. Its neighborhood characteristics, tree canopy coverage, and median household income are examined to determine how each factor can impact surface heat. Thermal energy data from NASA's Landsat 8 satellite were analyzed using QGIS, creating a high-resolution image that was then overlaid with census and tree canopy data.

Four different places were explored: one newer neighborhood, two older ones, and Downtown Lansing. All three neighborhoods were in the same school district, which minimized confounding factors. The findings of this study show that older neighborhoods had lower tree coverage and higher surface temperatures, showing that rather than income, neighborhood design and vegetation influence heat exposure. The study also discusses federal guidelines for urban heat mitigation as context for potential interventions. These results indicate that vegetation and urban design can outweigh income in shaping local heat exposure. This study is replicable in other mid-sized U.S. cities, providing a framework for implementing environmental science, policies, and urban planning to enhance a person's quality of life and urban heat resilience.

2. Introduction

Urban heat islands (UHIs) typically occur in areas with low presence of vegetation and high presence of heat-trapping surfaces. These surfaces could include concrete and other dense materials with low reflectivity. Rising temperatures cause heat-related illnesses, deaths, increased demand for cooling, and poor air quality, all of which can unevenly affect vulnerable populations. As the climate changes, the duration and intensity of heat events become more unpredictable. Therefore, understanding the cause of urban heat and working towards reducing it should become one of the main priorities for city planners and legislators.

Focusing on larger cities for UHI research has become common over the past several years. However, mid-sized cities, especially those in the Midwest, remain underrepresented in published papers, despite the fact that millions reside in these smaller areas.

Lansing can be used to represent many of these overlooked towns because it shares many characteristics with other Midwestern and post-industrial urban areas. Some examples include moderate populations and differences in tree canopy across block groups. Studying Lansing tests whether commonly cited urban heat factors hold when considering tree canopy and thermal imagery. Although these findings don't apply to all U.S. cities, they can still be used as a framework.

Furthermore, this study investigates changes in land temperature across the four selected areas. When observing heat differences, neighborhood age, tree canopy coverage, and median household income are all taken into account. By using satellite thermal imagery, census-level

socioeconomic data, and tree canopy estimates, urban heat mitigation is evaluated more deeply, rather than only highlighting the interactions that shape heat exposure. The study further investigates federal heat mitigation protocols to see how current policies help shape local conditions.

3. Methodology

3.1 Data Acquisition

Thermal data from May to June 2025 was obtained from NASA's Landsat 8 satellite. Landsat 8 can be considered a trustworthy source for urban climate studies due to its reliable data collection and level of detail. The imagery provided raw surface temperature values across most of Michigan, capturing spatial variations in heat distribution.

Using data from Census Reporter, socioeconomic data, such as median household income, shows data from the U.S. Census Bureau at the block group and tract levels. To be as detailed as possible, block groups were used. Block groups, the smallest geographic units within census tracts, were used for more detailed imagery. Tree canopy coverage estimates were found using the Global Forest Watch and provided approximate percentages of vegetative cover for each location.

3.2 Data Processing and Visualization

The thermal imagery from NASA was imported into QGIS, a software used to edit and analyze spatial data. QGIS was incorporated into this study because the data from NASA was shown in black and white. In QGIS, pseudocolor symbology was applied to show the data in a blue-to-red gradient. This addition enhanced the data and made the interpretation clearer.

Google Earth Pro was then used as a visual reference to locate Lansing's city boundaries and locate specific neighborhoods and landmarks. Representative points were then placed, typically near or in the locations.

Neighborhoods were selected to represent differences in development history while also minimizing confounding factors. All three neighborhoods fell within the same school district- the Lansing school district - to reduce variation linked to median household income and school district.

3.3 Location Selection

For this study, four locations were analyzed:

- Cherry Hill (newer area)
- Old Town
- Westside
- Downtown Lansing

Cherry Hill is one of Lansing's oldest neighborhoods, but has recently undergone a \$3 million renovation. For this reason, the newer residential area was analyzed rather than including

parts of its historic houses. Old Town and Westside represent older neighborhoods, and only the older areas were analyzed. Downtown Lansing was included to represent a complex urban environment because of the mixed land use, high surface coverage, and notable architectural features.

Since selecting entire neighborhoods and all of Downtown on Google Earth Pro is unrealistic, representative locations were chosen using well-known landmarks in or near the spot.

Table 1

Location	Cherry Hill	Westside	Old Town	Downtown Lansing
Category (New/Old/Complex)	New	Old	Old	Complex
Location Used	609 Cherry Street	Letts Community Center	Turner Dodge House	Michigan State Capitol Building
Median Income	\$33,045 ± \$21,562	\$81,838 ± \$32,206	\$52,137 ± \$40,379	N/A (Irrelevant)
Global Tree Watch Estimate	Roughly 35%	Roughly 20%	Roughly 20%	N/A (Irrelevant)
Block group, Tract	BG 1, Tract 67, Ingham, MI	BG 2, Tract 67, Ingham, MI	BG 4, Tract 8, Ingham, MI	N/A (Irrelevant)

Table 1 summarizes the representative locations, census tracts, block groups, and data sources used for each neighborhood.

4. Results

4.1 Thermal Patterns Across Neighborhoods

Thermal imagery revealed clear spatial variation in surface temperature across Lansing. Older neighborhoods had warmer surface temperatures compared to the newer neighborhood analyzed. As for Downtown Lansing, depending on the area, it had different surface temperatures. Areas west of the Capitol Building were warmer because of the dense building space, while the Capitol Building and areas east of it were cooler. The cooler temperatures were most likely a result of open green spaces and reduced spatial density. This highlights the influence of architecture and how it can increase or decrease UHIs.

Figure 2



Figure 2 is an image of Lansing using NASA's Landsat thermal imagery. Locations in or near the locations were mapped. 609 Cherry St: Cherry Hill. Turner-Dodge House: Old Town. Let's Community Center: Westside. Although the temperature changes seem small, they're significant during heatwaves, like the ones this past summer. Old Town and Westside have visibly higher surface temperatures than Cherry Hill, highlighting the role of tree canopy coverage.

Figure 3

Location	Tree Canopy %	Relative Heat Score 1→ Coldest 5→ Warmest
Cherry Hill	35%	2
Westside	20%	3
Old Town	20%	3
Downtown Lansing	N/A	4

Surface temperature scores were assigned on a 1–5 scale based on visual intensity from pseudocolor thermal imagery, where 1 = coolest and 5 = warmest.

4.2 Tree Canopy Data

The tree canopy coverage percentage varied between neighborhoods. Cherry Hill had a tree canopy coverage of approximately 30%. In contrast, Old Town and Westside had lower percentages at approximately 20%.

A pattern noticed was that locations with higher canopy coverage also had cooler surface temperatures. This reinforces the idea that vegetation can mitigate urban heat by providing shade

and evapotranspiration. Neighborhoods with higher tree coverage tended to be cooler, suggesting a negative relationship between vegetation and surface temperature.

4.3 Socioeconomic Data

Median household income differed across neighborhoods, with older neighborhoods displaying higher average income values than the newer neighborhood. However, income estimates had large margins of error due to census sampling limitations. Also, the data was used from the 2023 5-year estimates.

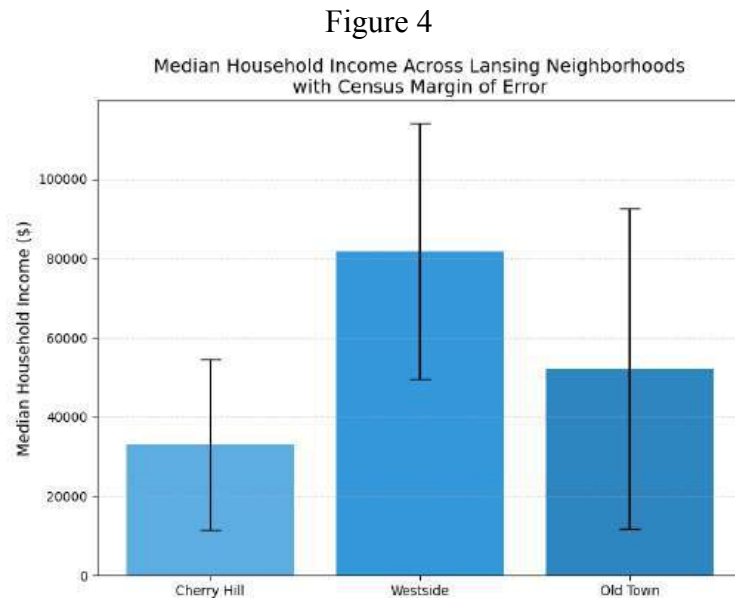


Figure 4 displays a bar chart showing the median household income across Lansing neighborhoods, as well as error bars representing the margin of error. Since the error bars overlap, the socioeconomic data is not statistically significant; therefore, due to chance.

Discussion

Older neighborhoods were observed to have higher median incomes and surface temperatures than newer neighborhoods. This suggests that when determining local heat exposure, neighborhood design, building density, and tree coverage can outweigh socioeconomic status. The lack of statistical significance in income differences doesn't weaken this study. Instead, it highlights the importance of avoiding oversimplified assumptions about wealth and environmental benefits. Heat exposure is shaped by architecture, the amount of vegetation, and policies over a period of time.

Historical development patterns may have a stronger influence than income on a neighborhood scale. Older neighborhoods are typically built before modern zoning regulations, so the lack of rules can lead to higher surface coverage and lower vegetation. In contrast, newer developments may have an advantage regarding contemporary planning practices, which incorporate tree planning and open space.

As for the complexity of Downtown Lansing, the Michigan state capitol represents a diverse thermal environment. While the Capitol building and its surrounding areas appear cooler, surface temperatures vary between places. As shown in Figure 2, these patterns are visible in

thermal imagery across Lansing. Areas with dense buildings and limited open space trap more heat, suggesting that mitigation strategies like reflective surfaces, green roofs, and targeted open spaces must be tailored to each location's layout. East of the Capitol, open spaces such as Grand River, Jackson Field, home of the Lansing Lugnuts Baseball Team, and the new Stadium district contribute to lower observed temperatures. In contrast, east of the Capitol has dense areas, which include apartment buildings, offices, and older structures with a low amount of open space and a high amount of pavement. That increases surface temperatures. These results suggest that tree planting, reflective surfaces, and open space planning should be tailored to neighborhood design rather than socioeconomic factors alone.

Policy Context and Implications

Several federal agencies have identified urban areas as a public health priority. To name a few, the CDC, EPA, and HUD have recommended tree planting, building structures with reflective materials, and an increase in green infrastructure to reduce UHIs. While these methods have been studied, this study suggests that their effectiveness depends mostly on conditions specific to neighborhoods.

Lansing illustrates how local conditions impact the effectiveness of specific mitigation strategies. Dense areas with open space, such as the Capitol grounds, show that vegetation and design can reduce surface temperatures. Nevertheless, older neighborhoods with limited space require alternative approaches. These findings suggest that a combination of historical development patterns, neighborhood layout, and vegetation coverage shapes urban heat vulnerability. While tree-planting can be effective in some areas, mitigation strategies should be tailored to local conditions.

Limitations

This research paper relied on relative surface temperatures instead of exact temperature values. In turn, it limits the ability to tell the exact thermal differences. Tree canopy estimates from Global Forest Watch are approximate and may not capture fine-scale vegetation variation. Additionally, census income data may contain uncertainty, which could be attributed to sampling error.

Even with these restrictions, the study's methodology can still be replicable for comparative urban heat analysis.

Conclusion

Heat exposure in mid-sized cities is driven more by physical characteristics of the built environment than by socioeconomic status alone. In Lansing, surface temperature patterns were most closely associated with neighborhood age, vegetation coverage, and land-use design. Although this study is primarily based on relative temperature and approximate vegetation data, it can still serve as a replicable framework for similar areas. It also underscores the significance of urban design in devising effective mitigation strategies. Future studies could incorporate in-situ temperature sensors or higher-resolution vegetation data to refine these findings.

Works Cited

- “Adapting to Heat.” U.S. Environmental Protection Agency | US EPA. Last modified on 19 March, 2025.
[https://www.epa.gov/heatislands/adapting-heat#:~:text=Provide%20community%20cooling%20centers%2C%20particularly,pdf\)%20for%20more%20information\).](https://www.epa.gov/heatislands/adapting-heat#:~:text=Provide%20community%20cooling%20centers%2C%20particularly,pdf)%20for%20more%20information).)
- Berube, Alan. “Small and midsize legacy communities: Trends, assets, and principles for action.” Brookings. Accessed January 3, 2026
<https://www.brookings.edu/articles/small-and-midsized-legacy-communities-trends-assets-and-principles-for-action/#:~:text=For%20more%20than%20a%20decade,to%20their%20businesses%20and%20residents.>
- “CDC Announces Important Advances in Protecting Americans from Heat.” Centers for Disease Control and Prevention | CDC (.gov). Accessed January 1, 2026.
<https://www.cdc.gov/media/releases/2024/p0422-heat-protection.html>
- “Census Reporter: Making Census Data Easy to Use.” Census Reporter. Accessed December 17, 2025. <https://censusreporter.org/>
- “City of Lansing Tree Canopy Report.” ArcGIS Story Maps. Accessed September 21, 2025.
<https://storymaps.arcgis.com/stories/2a02161311864252bd987582503ec319>
- Communications and Publishing. “The EarthWord: Evapotranspiration.” USGS (.gov). Accessed January 13, 2026. <https://www.usgs.gov/news/earthword-evapotranspiration>
- Druckenmiller, Hannah. “Urban Heat Islands 101.” Resources for the Future. Accessed January 3, 2026. <https://www.rff.org/publications/explainers/urban-heat-islands-101/>
- Hnavarro. “Analysis of urban heat islands identifies neighborhoods that expose 41 million residents to significantly higher temperatures.” Washington University in St. Louis. Accessed January 3, 2026.
<https://midwestclimatecollaborative.wustl.edu/cc-heat-island-analysis/#:~:text=News-,Analysis%20of%20urban%20heat%20islands%20identifies%20neighborhoods%20that%20expose%2041,illness%20and%20higher%20cooling%20costs.>
- Hollingsworth, Torey. “Revitalizing America’s Smaller Legacy Cities. Strategies for Postindustrial Success from Gary to Lowell” Lincoln Institute of Land Policy. Accessed September 22, 2025.
<https://www.lincolninst.edu/publications/articles/revitalizing-americas-smaller-legacy-cities/>
- Jiao, Min. “Optimizing the shade potential of trees by accounting for landscape context.” ScienceDirect. Accessed December 25, 2025.
<https://www.sciencedirect.com/science/article/abs/pii/S2210670721001931>
- “Lansing receives funding to add more trees in disadvantaged communities.” WKAR News. Accessed August 20, 2025.
<https://www.wkar.org/wkar-news/2024-10-13/lansing-receives-funding-to-add-more-trees-to-disadvantaged-communities>
- “Summer’s arrival was marked by a widespread June heatwave; dry conditions in the Northwest contrasted with wet conditions in the central and southern U.S.” National Centers for Environmental Information (NCEI) - NOAA. Accessed August 20, 2025.
<https://www.ncei.noaa.gov/news/national-climate-202506>

- “Urban heat islands and their adverse effects on their inhabitants.” Iberdrola. Accessed January 1, 2026. <https://www.iberdrola.com/sustainability/urban-heat-island>
- “What Are Heat Islands?” U.S. Environmental Protection Agency | US EPA. Last modified 2 September, 2025. <https://www.epa.gov/heatislands/what-are-heat-islands>
- “World Population Review.” World Population Review. Accessed January 2, 2026. <https://worldpopulationreview.com/us-cities/michigan/lansing#:~:text=Lansing%20is%20a%20city%20located,population%20of%20112%2C555%20in%202020>.
- “Your Gateway to NASA Earth Observation Data.” NASA Earthdata (.gov). Accessed August 19, 2025. <https://www.earthdata.nasa.gov/>

Harnessing the cBioPortal Database to Identify Genetic Determinants of Survival in Breast Cancer

By Claire Loftus

Abstract

Cancer is a complex and multifaceted disease characterized by genomic instability, unstable proliferation, evasion of growth suppressors, resistance to cell death, replicative immortality, angiogenesis, and the activation of invasion and metastasis. In fact, it is currently one of the leading causes of death worldwide. Hence, a better understanding of the causes, mechanisms, and behaviors of cancer is essential for creating improved prevention, detection, and treatment strategies.

Among the most frequent and deadly types of cancer is breast cancer. For women, the chance of developing breast cancer over one's lifetime is 13%, and the chance of dying as a result of breast cancer is 2.5%. Although these statistics remain concerning, the rise of genomic data promises a more optimistic future by offering invaluable insights into the underlying genetic mechanisms of breast cancer.

I utilized the cBioPortal database, a publicly available resource for exploring cancer genomic and patient data, to identify genes critical to the initiation and progression of breast cancer. More specifically, I analyzed the relationship between mRNA expression levels of specific genes linked to the cancer hallmarks—*CHEK1*, *BCL2*, *LEF1*, and *VEGFA*—and their respective impact on survival rates of breast cancer patients.

Upon investigation, I found that patients with high expression levels of *BCL2* and *LEF1* tend to survive longer, whereas patients with low expression levels of *CHEK1* and *VEGFA* tend to survive longer. These significant correlations indicate that the processes these genes are involved in may play an influential role in the advancement and survival of breast cancer.

Furthermore, these findings underscore the importance of databases, like cBioPortal, in making data more accessible to all. With greater access to research studies and clinical trial data, researchers and healthcare providers are able to collaborate, validate findings, and come to more generalized conclusions—thereby accelerating insightful discoveries.

Ultimately, cBioPortal facilitates ground-breaking cancer research and serves as a key component in the fight to end cancer as a whole.

Cancer

For several years, cancer has been a leading cause of death worldwide. Affecting approximately one in two individuals over their lifetime, with 32% not surviving beyond five years, cancer remains both highly prevalent and pernicious (“Cancer - NHS”)(“What’s Driving the Improvement”). Moreover, cancer does not discriminate; it impacts people of all ages, genders, ethnicities, and backgrounds.

Though great strides have been made towards treating cancer, it is still considered to be incurable. And of the treatments that do exist, many—such as chemotherapy and radiation—are often accompanied by severe side effects. For these reasons, identifying and comprehending trends within the data of cancer patients is essential for developing enhanced prevention, detection, diagnosis, and treatment strategies.

Part of this effort involves returning to foundational concepts to better understand what cancer truly is. At the surface level, it begins as a minuscule genetic mutation that causes an abnormal cell to form. Once the cell becomes cancerous and is unable to function properly, it begins to proliferate uncontrollably and potentially metastasize to other areas of the body. As the number of cancer cells increases, they continue to evolve and develop resistance to treatments by acquiring additional mutations. It is after the cancer cells have accumulated and potentially spread that they wreak havoc on the body—stealing nutrients, disrupting normal organ function, causing inflammation, and upsetting homeostatic chemical balances. All of these factors collectively contribute to cancer’s convolution and incurability (“What is Cancer?”).

Breast Cancer

Cancer encompasses a wide range of forms, among which breast cancer is one. Even within breast cancer, there are a plethora of distinct subtypes—the most common of which are invasive ductal carcinoma (IDC), invasive lobular carcinoma (ILC), ductal carcinoma in situ (DCIS), and lobular carcinoma in situ (LCIS). Diving further, there are subtypes defined by receptor status, including ER-positive (ER+), PR-positive (PR+), HR-positive (HR+), HR-negative (HR-), and HER2-positive (HER2+) (“Breast Cancer”).

The disease originates in the cells of the breast but has the ability to metastasize throughout the body via the blood or lymphatic system; indeed, approximately 80% of cases are classified as invasive. Not only is it highly invasive, but also widespread. Breast cancer is the most common cancer amongst women, impacting one in eight women during their lifetime. However, it is worth noting that breast cancer can also affect men, with approximately 2,600 men in the U.S. diagnosed each year (“Breast Cancer”).

Among both male and female cases, research has found that there is no single cause of breast cancer. Instead, there are common risk factors such as old age (55 and above), being biologically female, family history, genetic mutations, smoking, alcohol consumption, exposure to radiation, obesity, and ethnicity, among others. Regarding racial demographics, caucasian women are at a heightened risk for developing breast cancer compared to those who are Black, Asian, or Hispanic. That said, Black women face higher mortality from breast cancer, in part because they are more commonly diagnosed with triple-negative breast cancer—a more aggressive form of the disease (“Breast Cancer”).

In order to reduce the risk of breast cancer, individuals should avoid controllable risk factors (smoking and alcohol consumption), while also taking further preventative measures like receiving physical examinations and mammograms to detect early warning signs. If warning signs are present, the following tests are used to determine malignancy and diagnose breast cancer: breast ultrasound, MRI, biopsy, and an immunohistochemistry or genetic test (“Breast Cancer”).

Based on these tests, the breast cancer is staged on a scale from zero to four, considering tumor type, size, and location. Lower stages correspond to less advanced disease, whereas higher stages correspond to greater progression (“Breast Cancer”).

Once the stage of breast cancer is established, clinicians can devise a personalized and optimal treatment strategy. Most commonly, it is treated through surgery, chemotherapy,

radiation, immunotherapy, and targeted therapy. While these therapies may be successful, they are frequently accompanied by severe side effects, including fatigue, nausea, vomiting, and other gastrointestinal issues. Additionally, breast cancer surgery carries its own set of potential complications, such as infection, blood clots, nerve damage, and lymphedema (“Breast Cancer”).

It is thanks to these current treatments, coupled with advancements in the field that have led to earlier diagnosis and new therapies, that the future of breast cancer care appears brighter (“What’s Driving the Improvement”). For over thirty years, the breast cancer death rates have steadily been decreasing. Now, according to data maintained by the National Cancer Institute (NCI), the overall five-year relative survival rate for patients with breast cancer is 91%. When stratified by stage, the respective five-year relative survival rates for local, regional, and distant breast cancer are 99%, 86%, and 30% (“Breast Cancer”).

cBioPortal

The evolution of breast cancer treatments and the surge in survival rates would not be possible without help from databases such as cBioPortal. Originally created in 2008 by Memorial Sloan Kettering Cancer Center (MSK), cBioPortal is a publicly available database designed to facilitate the exploration of cancer genomics and patient data. Within this data, the platform allows users to search for genetic alterations such as mutations, copy number variants, and gene expression levels across a range of patients and types of cancer. Most importantly, cBioPortal makes genomic data accessible to all, including cancer researchers. Only through these datasets are researchers able to identify connections between genetic alterations and clinical outcomes, collaborate with other institutions, and make biological breakthroughs. At last, the findings are then implemented into routine clinical practice in order to improve patient care and outcomes (“cBioPortal”)(“Open Source”).

The Hallmarks of Cancer

When Douglas Hanahan and Robert Weinberg first proposed the concept in 2000, the *Hallmarks of Cancer* consisted of six defining capabilities: sustaining proliferative signaling, evading growth suppressors, activating invasion and metastasis, enabling replicative immortality, inducing angiogenesis, and resisting cell death. Over the ensuing decade, developments were made, and scientists’ understanding of cancer was broadened. Therefore, in 2011, the list was revised to include four additional capabilities: avoiding immune destruction, tumor-promoting inflammation, genome instability and mutation, and deregulating cellular energetics. Together, these ten hallmarks enable the formation and progression of neoplastic disease (Hanahan and Weinberg).

The discovery of such hallmarks revolutionized our interpretation of cancer in several ways. Crucially, it highlighted common themes among various types of cancer, thus facilitating the development of treatments applicable across multiple cancer types; targeted therapies are a prime example. Once the fundamental genes and mechanisms underlying each cancer hallmark were identified, targeted therapies were created with the aim of interfering with these specific processes. Since their introduction, targeted therapies have achieved an approximately 80%

success rate and are capable of targeting cancer cells while minimizing damage to surrounding healthy tissue (Pitakkitnukun).

Of the hallmarks described above, I chose to focus on four in particular: genome instability and mutation, resisting cell death, enabling replicative immortality, and inducing angiogenesis. The first, genome instability and mutation, is characteristic of most cancers because it refers to an increased rate of genome alterations such as insertions, deletions, point mutations, and genetic rearrangements. Genetic mutations, like these, many times play a role in the other hallmarks, including a cell's ability to resist death. Normally, apoptosis removes any unnecessary or unhealthy cells. Cancer cells, however, are capable of evading apoptosis by disrupting the balance between anti- and pro- apoptotic genes. Just as cancer cells can bypass apoptosis, they are also able to overcome the Hayflick limit—also known as enabling replicative immortality. In healthy cells, telomeres naturally shorten with each cell division and limit the cell's replicative capacity. Contrastingly, cancer cells tend to overexpress telomerase, which adds DNA to the telomere ends and allows for ceaseless proliferation. These cancer cells will continue to proliferate and grow as long as they have sufficient nutrients. It is through inducing angiogenesis, the stimulation of blood vessel growth, that cancer cells obtain a source of nourishment that provides the nutrients necessary for survival and metastasis. In sum, these four hallmarks are some of the most pivotal to cancer, as they enable the disease to inflict extensive damage on the body (Sumner).

Genes

I chose to study four genes associated with each of these four hallmarks and believed to influence breast cancer survival: *CHEK1*, *BCL2*, *LEF1*, and *VEGFA*. Although some of these genes are mainly known for their role in a single hallmark, it is important to note that most genes contribute to multiple hallmarks.

To begin, the *CHEK1* (checkpoint kinase 1) is part of the cell cycle checkpoint and aids in monitoring the cell cycle for errors in the DNA. Upon identifying such errors, *CHEK1* is activated and delays cell cycle progression to allow for the DNA to be repaired; if the DNA damage is irreparable, the *CHEK1* gene will then initiate apoptosis (Zhang and Hunter). In some cases, the *BCL2* (B-cell lymphoma 2) gene functions in opposition to *CHEK1* by inhibiting apoptosis, primarily by preventing the release of cytochrome c and other pro-apoptotic signals from the mitochondria. Not only does the *BCL2* gene prevent the release of pro-apoptotic signals, but it also binds to and deactivates proteins such as Bax and Bak to ensure cell survival (Czabotar et al.). Similar to the *BCL2* gene guaranteeing that the cell survives indefinitely, the *LEF1* (lymphoid enhancer-binding factor 1) gene makes certain that the cell divides indefinitely. To elaborate, the *LEF1* gene encodes a transcription factor that functions within the Wnt/ β -catenin signaling pathway. When this pathway is activated, *LEF1* partners with β -catenin to activate genes involved in cell proliferation and differentiation. Accordingly, abnormal *LEF1* gene activity leads to the constitutive expression of genes that drive cell division (“LEF1”). In the same way that *LEF1* activates the Wnt/ β -catenin pathway to promote cell division, it can also trigger the expression of pro-angiogenic genes like *VEGFA* (vascular endothelial growth factor

A). This gene codes for the *VEGFA* protein, which advances angiogenesis, increases vascular permeability, and induces vasodilation by stimulating endothelial cell proliferation and survival (Shibuya).

Materials and Methods

I began my analysis by navigating the dataset section within the cBioPortal website, locating the breast cancer category, and selecting the study titled “Breast Cancer (METABRIC, Nature 2012 & Nat Common 2016.) Within the study, I clicked “Query By Gene,” chose “mRNA expression” as the genomic profile, and entered the desired gene symbol. Once the query had been submitted and processed, I proceeded to the “Download” tab and downloaded the data file named “mRNA expression z-scores relative to all samples (log microarray).” Likewise, in the “Comparison/Survival tab,” I downloaded the data from the subcategory labeled “Survival.” After pasting the patient and survival data into a spreadsheet, I organized it using the “VLOOKUP” function; this allows users to search for specific values in the original column and retrieve corresponding values from another column. Following this step, I sorted the patient cases by mRNA expression level z-scores from highest to lowest and added a column labeled “High or Low Expression.” Any patient with an mRNA expression z-score lower than -1 was deemed “low,” while any patient with an mRNA expression z-score greater than 1 was deemed “high.” I then categorized the patients, grouping those with “high” mRNA expression separately from those with “low” expression. With the patients divided accordingly, I performed a t-test to determine whether the difference in means between the “high” and “low” expression groups was statistically significant ($p < 0.05$). If the results were significant, I calculated the average survival time (in months) for patients with “high” and “low” mRNA expression. By employing this method, I effectively highlighted the distinction between the two groups (Cerami et al.)(Gao et al.)(de Bruijn et al.).

Results

Utilizing the cBioPortal database, patients’ gene expression and survival data were extracted from the “Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016)” study. Patients with gene expression above a z-score of 1 were categorized as “high” expression, and those below a z-score of 1 were categorized as “low” expression.

Comparison of survival times between patients with high and low *CHEK1* expression revealed that those with low expression tended to survive longer (Figure 1A and 1B). The average survival time was 124 months for the low-expression group and 62 months for the high-expression group ($p < 0.001$). A Kaplan-Meier curve was generated to visualize survival differences between patients with high and low gene expression, and the “Log-Rank” test indicated a statistically significant difference between the two groups.

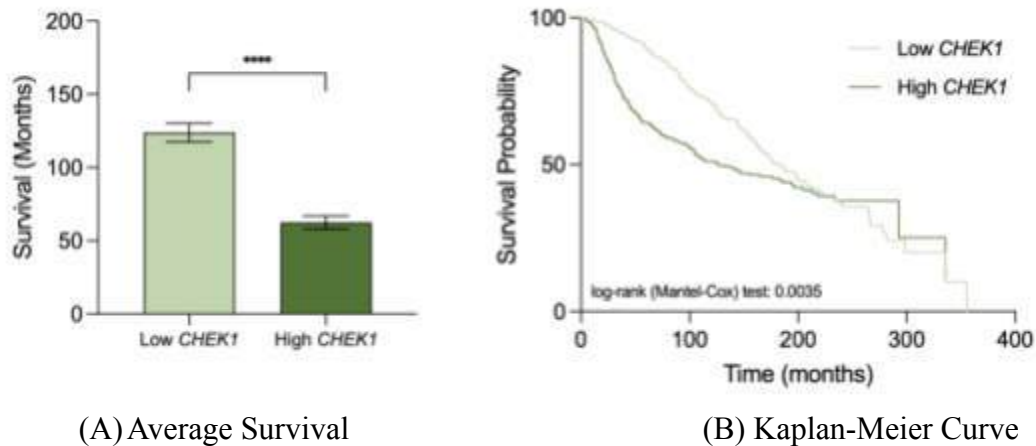


Figure 1. High *CHEK1* expression correlates with lower survival in breast cancer patients. Patients from “Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016)” were stratified into low or high *CHEK1* expression (see methods). (A) shows average survival and (B) shows a Kaplan-Meier curve. **** = $p < 0.0001$. Bars are mean $\pm$ standard error of the mean.

Comparison of survival times between patients with high and low *BCL2* expression revealed that those with high expression tended to survive longer (Figure 2A and 2B). The average survival time was 115 months for the high-expression group and 67 months for the low-expression group ($p < 0.001$). A Kaplan-Meier curve was generated to visualize survival differences between patients with high and low gene expression, and the “Log-Rank” test indicated a statistically significant difference between the two groups.

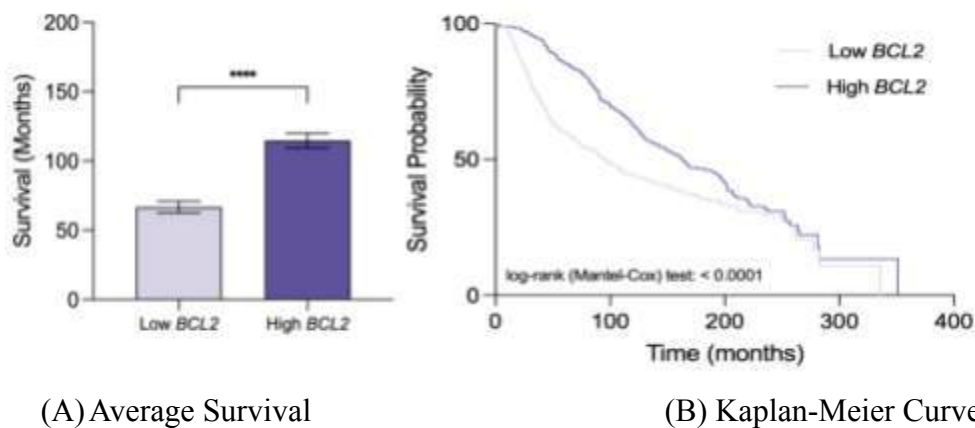


Figure 2. Low *BCL2* expression correlates with lower survival in breast cancer patients. Patients from “Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016)” were stratified into low or high *BCL2* expression (see methods). (A) shows average survival and (B) shows a Kaplan-Meier curve. **** = $p < 0.0001$. Bars are mean $\pm$ standard error of the mean.

Comparison of survival times between patients with high and low *LEF1* expression revealed that those with high expression tended to survive longer (Figure 3A and 3B). The

average survival time was 115 months for the high-expression group and 91 months for the low-expression group ($p < 0.001$). A Kaplan-Meier curve was generated to visualize survival differences between patients with high and low gene expression, and the “Log-Rank” test indicated a statistically significant difference between the two groups.

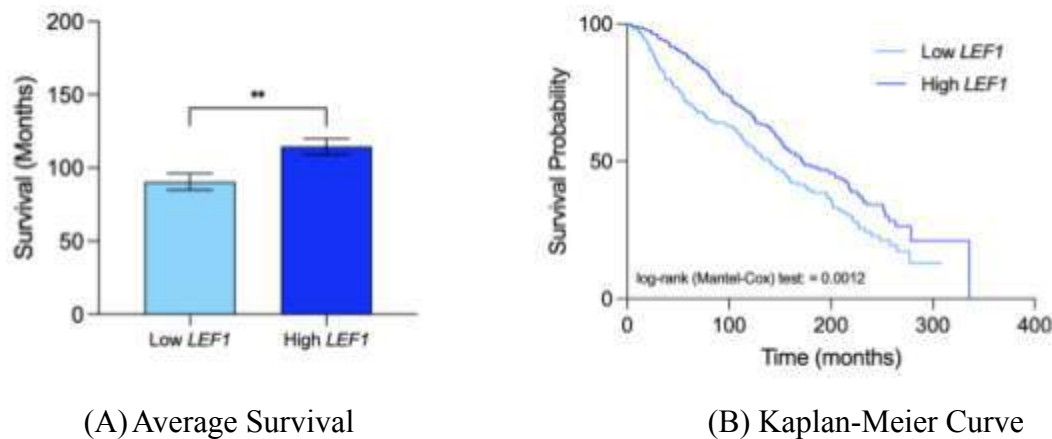


Figure 3. Low *LEF1* expression correlates with lower survival in breast cancer patients. Patients from “Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016)” were stratified into low or high *LEF1* expression (see methods). (A) shows average survival and (B) shows a Kaplan-Meier curve. ** = $p < 0.01$. Bars are mean $\pm$ standard error of the mean.

Comparison of survival times between patients with high and low *VEGFA* expression revealed that those with low expression tended to survive longer (Figure 4A and 4B). The average survival time was 102 months for the low-expression group and 76 months for the high-expression group ($p < 0.001$). A Kaplan-Meier curve was generated to visualize survival differences between patients with high and low gene expression, and the “Log-Rank” test indicated a statistically significant difference between the two groups.

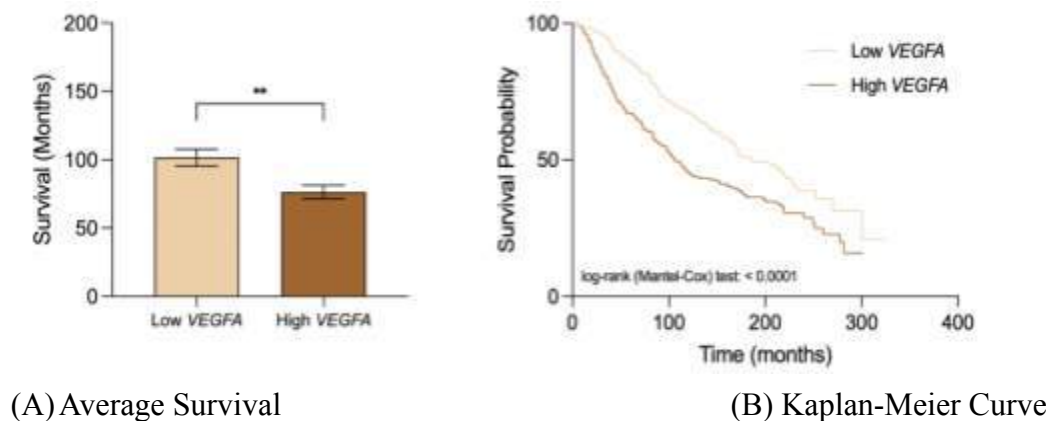


Figure 4. High *VEGFA* expression correlates with lower survival in breast cancer patients. Patients from “Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016)” were stratified into low or high *VEGFA* expression (see methods). (A) shows average survival and (B) shows a Kaplan-Meier curve. ** = $p < 0.01$. Bars are mean $\pm$ standard error of the mean.

Discussion

I explored the correlation between mRNA expression levels of genes and their respective effect on survival outcomes among breast cancer patients. The specific genes examined were selected based on their prominent roles in the hallmarks of cancer—namely, tumor-promoting inflammation, genome instability and mutation, resisting cell death, and inducing angiogenesis. Through my analysis of cBioPortal data, I determined that patients with elevated expression levels (z-score above 1) of *BCL2* and *LEF1* are inclined to survive longer, while patients with reduced expression levels (z-score below -1) of *CHEK1* and *VEGFA* are inclined to survive longer. My findings illustrate the notable interactions between the chosen genes and breast cancer progression and survival.

Given the known functions of each gene, most of the correlations identified appear to be plausible. Firstly, the *CHEK1* gene is heavily involved in the cell cycle checkpoint and the coordination of DNA repair. Underexpression of the *CHEK1* gene may allow cancer cells to progress through the cell cycle without adequate checkpoint control, eventually leading to genomic damage and tumor death. In contrast, elevated *CHEK1* expression may benefit the tumor by enabling cancer cells to repair DNA damage and successfully proceed through critical cell cycle checkpoints. Like *CHEK1*, my findings suggest that low *VEGFA* expression is also associated with prolonged patient survival. Seeing that *VEGFA* is key to angiogenesis, underexpression of this gene—and the resulting reduction in new blood vessel formation—would limit the tumor’s access to nutrients and its ability to survive.

Unlike *CHEK1* and *VEGFA*, overexpression of the *BCL2* gene can increase breast cancer patient survival. Although it may seem intuitive that lower expression of *BCL2*—a gene that regulates cell death by inhibiting apoptosis—would be advantageous by promoting tumor cell death, evidence suggests the opposite. The truth is that higher *BCL2* expression is linked to less aggressive tumors that are slow to progress and less likely to metastasize, so patient survival is slightly longer (Eom et al.). As with *BCL2*, elevated *LEF1* expression was associated with improved patient outcomes; however, this finding is less straightforward. Considering that *LEF1* promotes cell division, one might expect that its overexpression in cancer cells would indicate a more invasive, metastatic tumor and consequently shorten patient survival rather than prolong it. One possible explanation for this discrepancy is that the *LEF1* gene was overexpressed in anti-tumor immune cells. In such cases, increased replication of immune cells would strengthen the body’s immune response, enhancing its ability to combat cancer and ultimately lengthening breast cancer patient survival (Ho et. al.).

According to my results, it seems as though the dysregulation of the expression of *CHEK1*, *BCL2*, *LEF1*, and *VEGFA* may very well correlate with various processes involved in the initiation and development of breast cancer. To support this claim, I discovered several other sources that align with these results. For instance, in a 2020 study published by the *National Library of Medicine*, the authors state, “The probability of a reduced survival rate is greater for patients with high *CHEK1* gene expression in bladder, brain, lung, ovary, and breast cancers as compared with patients with low *CHEK1* expression” (Fadaka et al.). It is evident that, in breast cancer patients as well as across other cancers, patients with low *CHEK1* expression consistently

demonstrate longer survival as opposed to patients with high *CHEK1* expression. Another reputable source, the Multidisciplinary Digital Publishing Institute (MDPI), published an article where it was found that “Increased expression of Bcl-2 was associated with a protective effect on breast cancer-specific survival” (Al-Alem et al.). That is to say, amplified levels of *BCL2* expression correlate with better survival outcomes among breast cancer patients. Greater gene expression resulting in longer survival was also supported by a study published in *Anticancer Research*, claiming that “VEGF-A is a regulator of physiological and pathological angiogenesis whereby high levels of cytosolic VEGF-A correlate with shorter survival time in breast cancer” (Schmidt et al.). Simply put, they found that greater levels of *VEGFA* within cells lead to shorter survival times. Notwithstanding the identification of supporting literature for three of the four genes analyzed, *LEF1* represents a notable exception. Rather than upholding the idea that *LEF1* overexpression is favorable, much of the existing research reports the opposite effect. As previously mentioned, a potential rationale is that *LEF1* was overexpressed not in the cancer cells themselves, but in the surrounding, healthy immune cells. Taken together, current literature largely backs my findings.

Despite seeing strong correlations between expression levels of certain genes and survival for breast cancer, it is noteworthy that the same may not be true for other types of cancer; there are a plethora of factors that could be attributed to a lack of correlation. Fundamentally, it is likely due to the fact that every individual is biologically unique. Elements such as age, overall health, comorbidities, immune response, and cancer subtype vary greatly between patients, resulting in diverse outcomes. Moreover, although a specific genetic mutation may be correlated with survival in some cases, in others the correlation may be more complicated and require more rigorous stratification to uncover any significant associations.

For future analyses, I hope to harness the power of the cBioPortal database in alternative ways. One approach is to further stratify breast cancer patients into subgroups based on shared characteristics (e.g. estrogen receptor status or hormone receptor expression), thereby making it easier to understand how specific factors contribute to different clinical outcomes. Another approach is to provide a broader perspective on cancer by expanding the analysis to include additional studies and a wider range of cancer types. Regardless of the next steps taken, it is important to recognize that cBioPortal is more than a database that increases cancer data accessibility and facilitates analysis; rather, it serves as a stepping stone toward a better future for all.

Works Cited

- Al-Alem, Umaima, et al. "Prognostic Value of SGK1 and Bcl-2 in Invasive Breast Cancer." *Cancers*, MDPI, 11 June 2023, <https://www.mdpi.com/2072-6694/15/12/3151#:~:text=Increased%20expression%20of%20Bcl%2D2,%25%20CI%200.16%2C%200.65>. Accessed 10 Jan. 2026.
- "Breast Cancer." Cleveland Clinic, 25 Sept. 2023, <https://my.clevelandclinic.org/health/diseases/3986-breast-cancer>. Accessed 1 Jan. 2026.
- "Cancer - NHS." National Health Service, NHS Digital, 13 Oct. 2022, <https://www.nhs.uk/conditions/cancer/#:~:text=The%20cancerous%20cells%20can%20invade,of%20cancer%20during%20their%20lifetime>. Accessed 12 Jan. 2026.
- "cBioPortal." UCSF Helen Diller Family Comprehensive Cancer Center, University of California Regents, <https://cancer.ucsf.edu/research/molecular-oncology/cbioportal#:~:text=What%20is%20cBioPortal%3F,for%20researchers%20throughout%20the%20world>. Accessed 4 Jan. 2026.
- Cerami, E., et al. "The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data." *Cancer Discovery*, 2 May 2012, p. 401, <https://pubmed.ncbi.nlm.nih.gov/22588877/>. Accessed 4 Jan. 2026.
- Czabotar, Peter E., et al. "Control of Apoptosis by the BCL-2 Protein Family: Implications for Physiology and Therapy." *Nature Reviews Molecular Cell Biology*, Nature Publishing Group, 20 Dec. 2013, <https://www.nature.com/articles/nrm3722>. Accessed 4 Jan. 2026.
- de Bruijn, Elise, et al. "Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal." *Cancer Research*, 1 Dec. 2023, <https://pubmed.ncbi.nlm.nih.gov/37668528/>. Accessed 7 Jan. 2026.
- Eom, Yong Hwa, et al. "BCL2 as a Subtype-Specific Prognostic Marker for Breast Cancer." *Journal of Breast Cancer*, vol. 19, no. 3, 23 Sept. 2016, pp. 252–260, PubMed Central (PMC), <https://pmc.ncbi.nlm.nih.gov/articles/PMC5053309/>. Accessed 10 Jan. 2026.
- Fadaka, Adewale Oluwaseun, et al. "Gene Expression Alterations and Molecular Analysis of CHEK1 in Solid Tumors." *Cancers*, U.S. National Library of Medicine, 12 Mar. 2020, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7139733/>. Accessed 10 Jan. 2026.
- Gao, Jianjiong, et al. "Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal." *Science Signaling*, vol. 6, 2 Apr. 2013, <https://pubmed.ncbi.nlm.nih.gov/23550210/>. Accessed 3 Jan. 2026.
- Hanahan, Douglas, and Robert A. Weinberg. "Hallmarks of Cancer: The Next Generation." *Cell*, vol. 144, no. 5, Elsevier B.V., 4 Mar. 2011, pp. 646–674, https://www.sciencedirect.com/science/article/pii/S0092867411001279?ref=pdf_download&fr=RR-2&rr=7f9532d2480626bb#fig1. Accessed 9 Jan. 2026.
- Ho, Ngai, et al. "LEF1 Drives a Central Memory Program and Supports Antitumor Activity of Natural Killer T Cells." *Cancer Immunology Research*, vol. 11, no. 2, 3 Feb. 2023, pp. 171–183, PubMed Central (PMC), <https://pmc.ncbi.nlm.nih.gov/articles/PMC9898189/>. Accessed 8 Jan. 2026.

- “LEF1.” Online Mendelian Inheritance in Man (OMIM), McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 153245, OMIM.org, <https://omim.org/entry/153245>. Accessed 3 Jan. 2026.
- “Open Source: MSK Researchers Share Valuable Cancer Data.” Memorial Sloan Kettering Cancer Center, 20 May 2021, <https://www.mskcc.org/news/open-source-msk-researchers-share-valuable-cancer-data#:~:text=Sharing%20information%20to%20expedite%20discoveries,cBioPortal%20was%20launched%20in%202008>. Accessed 12 Jan. 2026.
- Pitakkitnukun, Phuchai. “Targeted Therapy: Stopping Cancer in Its Tracks.” Samitivej, 5 Mar. 2018, <https://www.samitivejhospitals.com/article/detail/stopping-cancer-targeted-therapy#>. Accessed 8 Jan. 2026.
- Schmidt, Melanie, et al. “Expression of VEGFR-1 (Flt-1) in Breast Cancer Is Associated with VEGF Expression and with Node-negative Tumour Stage.” Anticancer Research, The International Institute of Anticancer Research, 26 Mar. 2008, <https://ar.iijournals.org/content/anticancer/28/3A/1719.full.pdf>. Accessed 2 Jan. 2026.
- Shibuya, Masabumi. “Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signaling in Angiogenesis: A Crucial Target for Anti- and Pro-Angiogenic Therapies.” Genes & Cancer, U.S. National Library of Medicine, Dec. 2011, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3411125/#:~:text=VEGF%2DA%20binds%20to%20and,cell%20migration%2C%20and%20gene%20expression.&text=In%20addition%2C%20Lee%20et%20al,cells%2C%20contributing%20to%20endothelial%20functions>. Accessed 1 Jan. 2026.
- Sumner, Chris. “The Guide to the Hallmarks of Cancer Research Targets.” Cell Signaling Technology, https://blog.cellsignal.com/the-guide-to-hallmarks-of-cancer-research-targets?_gl=1*_unxth0*_ga*MTc0MjI5ODMxNS4xNjk3ODY0MzQ2*_ga_L0Q98CEH39*MTY5Nzg2NDM0Ni4xLjEuMTY5Nzg2NDk0MS41OC4wLjA. Accessed 14 Jan. 2026.
- “What is Cancer?” National Cancer Institute, 11 Oct. 2021, [https://www.cancer.gov/about-cancer/understanding/what-is-cancer#:~:text=When%20cells%20grow%20old%20or,or%20not%20cancerous%20\(benign\)](https://www.cancer.gov/about-cancer/understanding/what-is-cancer#:~:text=When%20cells%20grow%20old%20or,or%20not%20cancerous%20(benign)). Accessed 14 Jan. 2026.
- “What’s Driving the Improvement in U.S. Cancer Survival Rates?” City of Hope, 26 Jan. 2023, <https://www.cancercenter.com/community/blog/2023/01/cancer-survival-rates-are-improving#:~:text=Five%2Dyear%20survival%20rates%20have,currentl%20sits%20at%2068%20percent>. Accessed 10 Jan. 2026.
- Zhang, Youwei, and Tony Hunter. “Roles of Chk1 in Cell Biology and Cancer Therapy.” International Journal of Cancer, U.S. National Library of Medicine, 1 Mar. 2014, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3852170/#:~:text=While%20the%20major%20function%20of,viability%2020%2C%2075%E2%80%93377>. Accessed 6 Jan. 2026.

Consumed By Perception: Evaluating How Glorified Diets and Superfoods Reshape Societal Norms By Sydney Pryor

Societal glorification of food is growing more common in modern society. Focusing on the extent this glorification can affect society is vital to understand. As social media continues to grow, connections to internal societies also increase. These connections in society can contribute to how we view different cultures. This research explores why society often focuses on past diets over modern ones, and how this focus impacts society as a whole- specifically in shaping future generations.

It is vital to use philosophy and scientific inquiry to make strong connections between society and ancient diets. Many scientists in past generations spoke about diets frequently to the general public. Albert Einstein, for example, spoke about the diet of vegetarianism as early as 1954, which likely cascaded into the glorification of that specific diet. “Nothing will benefit human health and increase chances of survival of life on earth as much as the evolution to a vegetarian diet” (Einstein).

Establishing diet ethics within a wide range of societies creates an important connection between these topics and comes to a conclusion as to why these topics are so easily connected. Through a cultural and ancestral lens, food shapes both identity and evolving societal paths. Through a generational lens, it is easily highlighted how different populations, both older and younger, interact with diet trends in a different manner which can very likely lead to the challenging of dietary norms. These diets are assumed to lead to high health for humans but there is debate about what makes a diet healthy “There is general agreement that healthy and high-quality diets should provide sufficient energy and all the essential nutrients and non-nutritive factors, like fiber, that people require” (Lesica 1).

The psychological appeal to this topic of glorifying diets is prominent in the health industry as it relates to the societal view that food is either an enemy or an all-healing medicine. One of the objectives of this paper is primarily focused on the extent to which diets and superfoods affect modern and past societies, while also looking at the impact through Cultures, Generations, Ancestral, and Psychological Factors.

Generational

All across the world, diets have been altered through generations, though in certain modern societies, older diets like Mediterranean diets are being glorified because they seem to have improved preceding generations' health and body image. There are many differences across society in different age cohorts. A large factor for the younger generation in this glorification of diets is primarily to obtain a better-looking body. Many younger individuals focus on diets outside of their culture because it makes others look healthy. “Luana follows a strict diet recommended by a nutritionist in Lecce, and her mother now also enjoys eating the raw greens, seeds, and dry salads that Luana discovered on a recent trip to India” (Nicolescu 121). Following strict diets found abroad that seem “healthy” alters the way society looks upon their cultural food creating a sense of dishonor and replacement for the food they created and instead opting for the food that will give them the “perfect body”.

As we move away from why individuals follow diet trends from other countries, we look at other factors like what “diet ethics” are as a whole. Diet ethics refers to the moral principles and values that create a guide for our food choices and dietary practices personally. Many healthcare professionals deal with diet ethics, if generations of individuals have poor diet ethics then it is likely that these poor dietary habits will continue. “Apart from pointing out the harm caused by bad healthcare, these references sometimes also highlight the shame felt when one can no longer conceal these diseases” (Van Hoof 124). Modern-day societies view food as something that can harm oneself and create a poor image though others, who look up to food as a necessary and helpful subject, do not fear food itself but rather the food that is not nourishing. “One should enjoy eating either necessary or pleasant foods; when not hungry, one should not create any extra appetites” (Roskam & Van der Stockt 119).

American societies often glorify eating constantly, while eating should be something that we do when we are hungry or just want something to satisfy a craving. Compulsive eating habits prevent future generations from developing a healthy relationship with food, and create a negative image in America’s diet culture. To fully understand this trend, we must also examine the psychological motivations that reinforce glorified eating habits across all generations.

Psychological

By looking at and researching how generations interact and view these different ancient diets from other countries, another impact that must be investigated are the psychological factors. These factors can persuade individuals, through all generations, to glorify diets. Examining how diets affect individuals who struggle mentally, can create a well-rounded and well-understood statistical analysis. Psychotherapy is the diagnosis and treatment of mental and emotional disorders. Within the realm of psychotherapy, many concepts tie into this form of therapy. For example, schizophrenic and autistic children, as well as children with learning difficulties, have reportedly been treated successfully with nutrition therapy (Cott, 1971; Rimland, 1968). This piece of evidence supports the idea that diets can effectively contribute to a mentally-different patient's overall demeanor. Creating a sense that diet is highly likely to be tied to the mental health of those who are struggling psychologically or emotionally.

Diets can also affect one's ability to aid the way families of those with mental health issues or disorders view diets. “Critics of consumerism often see marketing as the process of instilling new desires, but this manual reveals that the secret of children’s mass consumption is instead teaching children to reject often-familiar things they have already learned to like at home, to acquire a complex set of distastes”(Wilk 18). This alters one's view from being strictly to use diets to make someone “look picture perfect” to being beneficial to one's mental health.

As this research continues down the path of healthcare and how diets can directly associate with the society within these healthcare systems, it easily notices that there is a parallel to the glorification of diets and how psychological motivations occur in populations. Through some recent research papers like “ The Alzheimer's Associations Special Report, Mapping a Better Future for Dementia Care Navigations” it is highlighted that both patients and caregivers are part of the society that is directly affected by diets in the healthcare system. There is an urge to seek dietary trends to provide control, healing, and a sense of identity for patients and workers

as they go through health crises and uncertainty on a day-to-day basis. “Daily health care activities may include scheduling appointments with health care providers, attending doctor's visits, and scheduling social and community support for themselves and the person living with dementia, such as in-home assistance, adult day programs or meal delivery” (Alzheimer’s Association 94). This piece of evidence supports how diets within healthcare facilities can truly alter the way the overall feeling is among the hospital population. Without the inclusion of diets, meal delivery, and healthy foods the population would be directly affected.

Cultural

In addition to psychological factors, cultural norms are also significantly shaped by diets. In many countries, it is a taboo to consume certain foods in your diet, while in others it is seen as a tradition and honor to be eating that same food. This can be primarily based on religious aspects in a country, or could be based on the surrounding environment.

A specific article focuses on the diet shift due to the culture in the environment and how society viewed animals. “It is possible to adopt a diet that benefits the health of humans and reduces or avoids the suffering of animals, including animals killed for consumption, animals involved in agricultural research, and animals used for research and testing of drug interventions and diseases” (Vyas 127). Over time these taboos have turned into a dietary guideline for many to improve one's health. An article that supported my research, discussed pre-Columbian Huecoid and Saladoid cultures’ diets and their surrounding environments. Through this article, it highlighted many connections between the way of living and diets. Their diets were primarily based upon what was easily accessible and what was not going to send their way of living into “destruction” or “chaos”. Specifically, this pre-Colombian diet contained fish and smaller organisms like mollusks and crustaceans and some smaller mammals. This became a prominent diet in their culture and was not seen as non-uniform. This population dwindled significantly due to a parasite found in the diet of this area, later found due to parasite-infected hosts. However, after continuous research on these specimens, and extensive phylogenetic tracing, “four whole canid remains were found laid out in a burial disposition suggesting endearment towards an animal companion, while scattered canids osseous remains without anatomical context were also dispersed throughout the deposits, suggesting canids as an occasional food source” (Rosana Wiscovitch-Russo 9). The research behind this consumption showed human evidence of consumption on smaller mammals like canids though now it would likely be seen as a taboo hence why it is not common to see this consumption in many countries as there is a parasitic risk and stigma against it.

As many cultures diverge, and as time continues, there is always something tying the cultures together through time. Cultures develop food customs and dietary traditions. In the case of the Incas, there was concern over the agri-food system because most of the Incas’ food sources were agriculturally based due to the accessibility and convenience of the easily-grown food. “However, primitivist representations of superfoods essentialize producers and production practices as traditional and timeless, obscuring their complex and changing reality” (Loyer 2). This source of evidence aids in showing how superfoods are glorified to an extreme extent, meaning in a way that it is altering how society sees reality as a whole.

Ancestral

There are countless ways that dietary glorification of older foods can be passed down through cultural traditions, many of which are related to traditions rooted in identity or survival passed down through ancestors. Many articles have spoken about how family cuisine, and their family's traditions, affect the way families gather around the dinner table, and what foods they consume. "Greeks have been feasting on delicacies such as roasted lamb with capers, boiled or fried seafood, eggs, cheese, saffron-flavored rice, lentil or bean soups, honey cakes" (Flores 2).

As society as a whole grows, so does the way we preserve our past family. In a memoir called "My Mother's House," it ties in how, generation to generation, the way we live changes even if we still feel nostalgic. This nostalgia can affect how we preserve our modern-day lives and the way we live them. It links this nostalgia to how Haitian food, along with ritual, develops the adoption of ancestral knowledge. "A fig tree that fruited in August, and in September, sitting in the sun after school, still wearing her school clothes but no shoes, eating syrup-preserved figs on stale Melba toast that exploded in her cheeks as she sat on the front step of the house" (Ghansah 10). This quote connects how eating a simple piece of fruit can bring back that sense of nostalgia for a certain place and create an image in perspective of what used to be.

As time passes, more change is made between cultures and the way societies function. An article by John D Flores continues to talk about the "higher" cause within their culture. "The ancient Greeks battled countless times among themselves, but they always put their differences aside for a "higher" cause to honor their gods, to fight the Trojan war, to compete in the Olympics, to defend the motherland from Persian invasions, to support Alexander the Great in his noble conquest to transfer Greek thought, ideals and knowledge to the rest of the world" (Flores 2). This centrality creates that sense of family and continuous rituals over the dinner table, creating a sense of character for food and showing how diets in a culture shape the way society sees certain rituals. These diets, specifically Greeks, create a healthy and nourishing environment at a dinner table and a positive cultural ambiance reinforcing a sense of community and identity passed down through cultural legacy.

Conclusion

By evaluating the evidence through multiple lenses, including ancestral, generational, psychological, and cultural, it is clear that the glorification of diets and many superfoods significantly impacts society. These influences create alterations in how individuals view their health, traditions, and identity. Some of these diets are used as cultural taboos, or a sign to start changing social norms, while these superfoods and diets are also incorporated for aesthetic or healing purposes. Overall, glorified diets and superfoods serve as both a reflection of society and even a force that alters society in either positive or negative ways.

Works Cited

- Alzheimer's Association. Special Report: Mapping a Better Future for Dementia Care Navigation. Alzheimer's Association, 2024,
<https://www.alz.org/getmedia/a7390c29-e4c0-4d3c-b000-377b6c112eea/alzheimers-facts-and-figures-special-report.pdf>.
- Cook, Sarah. "Diverse, High-Quality Diets for Human Health." *The Spice of Life: The Fundamental Role of Diversity on the Farm and on the Plate*, International Institute for Environment and Development, 2018, pp. 19–26.
- Floros, John D. *Food and Diet in Greece From Ancient to Present Times*, 2004,
<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=bbe1c976f3652377073334f930f86db06d05339c>.
- Flores, John D. "Cultural Unity in Ancient Greece." 2023.
- Ghansah, Rachel Kaadzi. "My Mother's House." *Transition*, no. 109, 2012, pp. 3–19.
- Katzenstein, Leonard. *Einst: The Magazine for Alumni and Friends of Albert Einstein College of Medicine of Yeshiva University*, 2014,
<https://einsteinmed.edu/docs/publications/einstein-magazine/einstein-magazine-winter-spring-2014.pdf>.
- Loyer, Jade. "Selling the 'Inca Superfood': Nutritional Primitivism in Superfoods Books and Maca Marketing." *Gastronomica*, 2018,
<https://www.tandfonline.com/doi/abs/10.1080/15528014.2018.1480645>.
- Niculescu, Razvan. *The Imposition of Beauty*. 2016,
<https://www.jstor.org/stable/pdf/j.ctt1gxxpdq.9.pdf>.
- Roskam, Geert, and Luc Van der Stockt, editors. *Virtues for the People: Aspects of Plutarchan Ethics*. Leuven UP, 2011.
- Van Hoof, Lieve. "Plutarch's 'Diet-Ethics.'" *Virtues for the People*, edited by Geert Roskam and Luc Van der Stockt, Leuven UP, 2011, pp. 109–130.
- Vyas, Sheetal V. "Disease Prevention with a Plant-Based Lifestyle." *Animal Experimentation*, edited by Kathrin Herrmann and Kimberley Jayne, vol. 22, Brill, 2019, pp. 124–148.
- Wilk, Richard. "Loving People, Hating What They Eat." *Reimagining Marginalized Foods*, edited by Ellen Finniss, U of Arizona P, 2012, pp. 15–33.
- Wiscovitch-Russo, Rosana. "Pre-Columbian Zoonotic Enteric Parasites." 2020,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC6992007/pdf/pone.0227810.pdf>.

An AI-Powered Night Vision System for Pedestrian Detection and Illumination at Crosswalks Using a Custom YOLOv8 Model By Agustya Agustya

Abstract

Pedestrian and cyclist safety at urban intersections during low-light conditions remains a critical issue. As per the NHTSA, nearly 40% of all bicycle accidents occur at intersections. More than 50% of these accidents occur during low-light conditions. This paper introduces the Night Vision Crosswalk Guardian, an event-aware, privacy-preserving pedestrian detection and targeted illumination system that addresses the limitations of generic motion-activated lighting solutions by implementing class-gated detection (person-only) and crosswalk-specific deployment geometry. A custom-trained deep-learning-based object detection model was developed using Sagemcom's night vision computer vision project dataset and a pre-trained YOLOv8 model that achieved a test performance of 84.6% mAP50. In controlled low-light tests, the prototype detected pedestrians in 100% of trials where power supply was adequate, being able to detect subjects as soon as they entered the frame. Unlike typical surveillance AI, this system actively enhances road safety through real-time interventions. The prototype demonstrates a novel integration of deep-learning object detection with smart infrastructure for nighttime pedestrian protection.

Keywords

Deep Learning, Machine Learning, Night Vision Crosswalk Guardian, YOLOv8, Convolutional Neural Network (CNN), Object Detection, Transfer Learning

Introduction

Pedestrian and cyclist safety remains a critical public health concern, particularly in high traffic density areas and after dark when visibility is reduced. According to the National Highway Traffic Safety Administration (NHTSA)¹, nearly 40% of all bicycle accidents occur at intersections, and more than 50% happen during low-light conditions such as dawn, dusk, or nighttime. Limited visibility—due to insufficient street lighting, dark clothing, non-reflective gear, non-functional bike lights, and/or insufficiently bright streetlights—is a major contributing factor to these accidents.

Multiple studies have shown that proper lighting significantly improves nighttime safety. The FHWA Lighting Handbook notes that roadway lighting can reduce fatal nighttime crashes by up to 65% by improving visibility of pedestrians, cyclists, and other road users². The Pedestrian Lighting Primer similarly reports crash modification factors (CMFs) showing reductions in pedestrian crashes ranging from 42% to 81%, depending on crash severity³. A case study by the FHWA demonstrated that installing intersection lighting reduced late-night/early-morning crashes by an average of 35%⁴.

If we can keep a light source constantly focused on a cyclist or a pedestrian whenever they are crossing the street, it can potentially minimize the number of accidents that occur while people are crossing streets. A focused light source ensures the pedestrian or cyclist has caught

the attention of the driver, ensuring that they will slow down to avoid a collision with the crossing person. Also, a single moving light source can cover a large area, potentially saving energy.

To improve pedestrian and cyclist safety, cities around the world have tested a range of technologies—from smart crosswalks with embedded LED lights to infrared detection systems and AI-powered camera monitoring. Companies like FLIR Systems use thermal imaging to enhance nighttime visibility ⁵, while cities like Oslo and Helsinki have successfully reduced pedestrian fatalities through smart infrastructure and strict traffic enforcement. Products like the Lightguard Systems IRWL ⁶ and the Crosswalk Signals Smart Pedestrian Warning Bollard ⁷ offer roadway-embedded lights and illumination-based alerts. However, these solutions still fall short in certain conditions—especially in low light—where visibility remains a challenge. This is where the proposed solution comes in.

Our proposed solution is an AI-powered Night Vision Crosswalk Guardian. The device uses an infrared (IR) camera paired with a deep learning based object detection model designed to identify pedestrians and cyclists crossing the street in real time. Upon detection, the system activates an LED array as a substitute for a floodlight to illuminate the individual, providing a clear visual signal to approaching drivers.

Unlike conventional motion-activated lighting systems that trigger illumination for any movement (including vehicles), our proposed solution implements two critical gating mechanisms: class-gated detection that responds exclusively to pedestrians and cyclists through trained object recognition, and physically targeted coverage limited to crosswalk zones through strategic mounting and field-of-view configuration. This dual-gating approach transforms generic lighting into event-aware, safety-specific infrastructure that enhances driver alerting without contributing to light pollution or false activation noise.

To build the deep learning model, transfer learning (adapting a pre-trained model to a specific task) was applied to the most popular object detection model, YOLOv8 (You Only Look Once version 8, a state-of-the-art real-time object detection algorithm). 25 training sessions were conducted to identify the optimal configuration. Finally, the model was deployed on a Raspberry Pi with an IR camera and a light source to complete the solution.

After reviewing related work on pedestrian and cyclist safety technologies, including work done by other researchers in this field, the proposed AI-based solution is presented. The methods used are detailed, including dataset preparation, model architecture, training setup, and hardware/software configuration. The results section presents the YOLOv8 model performance with various hyperparameters. Finally, the implications of our findings, challenges faced during development, deployment considerations, and cost-benefit analysis are discussed, concluding with a summary and suggestions for future work.

Related works

Pedestrian detection in low-light and nighttime conditions has been extensively explored across various domains, including surveillance, autonomous vehicles, and safety infrastructure. Recent works in this area have focused on integrating thermal imaging, sensor fusion techniques,

deep learning models, and improved datasets to enhance pedestrian visibility and responsiveness in safety-critical environments.

Park et al.⁸ developed a sensor-fused nighttime system for pedestrian detection in Advanced Driver Assistance Systems (ADAS) and autonomous vehicles, combining thermal and RGB cameras with a novel image alignment algorithm to optimize fusion. They evaluated five deep neural network (DNN) models, identifying late fusion as optimal in balancing detection accuracy (95.5% mean Average Precision at 50% Intersection over Union threshold (mAP50)) and real-time inference speed (33 fps on Jetson Orin device). They also collected 32,000 nighttime image pairs, each pair containing one RGB and one infrared (IR) image of the same subject, to support robust model training. This approach suggests future improvements for this work, such as adding both RGB and IR cameras to enable detection during both day and night and to enhance accuracy.

Ghose et al.⁹ proposed augmenting thermal images with saliency maps to improve daytime pedestrian detection using only thermal inputs. They trained a Faster R-CNN with deep saliency maps generated using R3-Net, achieving a 13.4% and 19.4% reduction in miss rates for day and night images, respectively, compared to the baseline model. They also published pixel-level saliency annotations for a subset of the KAIST dataset, which could be valuable for model training. This work supports the idea that incorporating saliency maps in a real-world implementation of our device could further improve pedestrian detection. Additionally, we can leverage the KAIST dataset to significantly boost the accuracy and robustness of our deep-learning model.

Liu et al.¹⁰ introduced PTB-TIR, a benchmark dataset with 60 manually annotated thermal infrared (TIR) sequences for pedestrian tracking. The dataset includes nine attribute labels to test algorithm robustness under conditions such as occlusion, scale variation, and low resolution. Their comparative analysis of nine tracking algorithms across three tracker components—feature extractor, motion model, and observation model—showed that the feature extractor was the most critical factor influencing tracking performance.

Kulhandjian et al.¹¹ designed a multi-sensor pedestrian detection and alert system using visible, infrared, and radar sensors. Their system utilized deep convolutional neural networks (DCNNs) trained on multimodal data to detect pedestrians and alert drivers with 98% accuracy during both day and night scenarios. The hardware was deployed in real-time on a vehicle platform using a Raspberry Pi and a Coral deep-learning accelerator. This study highlights the effectiveness of sensor fusion and machine learning in real-time pedestrian detection and driver alert systems.

Geng¹² implemented pedestrian detection on infrared images using YOLOv3 combined with transfer learning and Distance Intersection-over-Union (DIoU) loss. Using the CVC dataset, the Diou-YOLOv3 model exhibited faster loss convergence but lower detection accuracy (72.14%) compared to standard YOLOv3 (96.35% AP). This experiment highlights the trade-offs between detection accuracy and model convergence in low-visibility conditions. The significant performance gains achieved using YOLOv8 in our research illustrate recent advancements in object detection algorithms.

Akilan and Vachhani ¹³ provided a comprehensive review of low-light pedestrian detection algorithms, identifying major challenges and techniques involving region-based (R-CNN), non-region-based (SSD, YOLO), and graph neural network (GNN) methods. They also reviewed major datasets, emphasizing the need for fusion techniques and robust models that operate reliably across diverse lighting conditions.

Nowosielski et al. ¹⁴ presented an embedded night-vision system for pedestrian detection using low-cost thermal imagery processed on an ODROID XU4 microcomputer. They compared classical detection algorithms (YOLOv2, HAAR + AdaBoost) and introduced an encoder-decoder DNN with attention mechanisms to enhance detection accuracy while ensuring energy efficiency and portability in embedded systems. Their work highlighted the feasibility of low-cost real-time pedestrian detection systems in dark environments.

Raza et al. ¹⁵ developed a pedestrian detection and tracking framework for nighttime surveillance using IR imagery. They employed random forest segmentation, multiple template matching, and particle filtering, achieving a detection accuracy of 90% and tracking accuracy of 81% using the OSU Thermal Pedestrian Database. Their system was particularly effective for static camera surveillance applications.

Wang et al. ¹⁶ introduced the NightSurveillance dataset, addressing the limitations of existing nighttime pedestrian datasets by focusing specifically on real-world surveillance scenarios. The dataset includes over 38,000 frames with diverse attributes such as occlusion, blur, and adverse weather. Their analysis showed that even state-of-the-art detectors trained on daytime datasets struggled in these conditions, emphasizing the importance of using specialized surveillance-focused datasets.

These prior studies demonstrate the value of thermal imaging, multi-sensor fusion, and saliency augmentation in nighttime pedestrian detection. They highlight the importance of specialized models, diverse datasets, and real-time embedded deployment strategies. These studies inform our research, which integrates a YOLOv8 model with a dataset of nighttime images and deploys it on an IR-camera-based Raspberry Pi system to improve pedestrian visibility at crosswalks during low-light conditions.

Materials and Methods

The overall research has 2 phases: first, building a deep learning model that can detect pedestrians, and second, building a device that uses the model to detect pedestrians on the camera feed and focus the light on them.

Dataset

The dataset in YOLOv8 format was obtained from Roboflow, a popular database for object detection models ¹⁷. The dataset contains 4000 images with labeled persons.

Deep learning model

To detect pedestrians, a custom deep learning-based object detection model was developed. The steps for developing the model are as follows.

Model Training

To train and build the model, the dataset was split into training (80%), validation (10%), and testing (10%). For the project, only the “person” label was needed, so code was used to remove all the images with no people in them. The Ultralytics Python library was used to download the pre-trained YOLOv8 model¹⁸. Then, transfer learning was performed with the dataset to create a custom model¹⁹. The model was trained across different epochs (training cycles) ranging from 10 to 50 and learning rates (the rate at which the model learns) between 0.01 to 0.0001. The model output from each experiment was validated using the validation dataset, and the results were recorded in a Google Sheet. The model with the highest mAP50 value was then tested with the test dataset.

Training Configuration and Hyperparameter Selection

Parameter	Range Tested	Selected Value	Justification
Learning Rate	0.0001-0.01	0.0005	Optimal convergence without overshooting
Training Epochs	10-50	50	Performance plateau observed after 40 epochs
Batch Size	-	16	Hardware memory constraints
Image Size	-	640×640	YOLOv8 default, inference speed balance
Optimizer	-	AdamW	Ultralytics default
Data Split	-	80/10/10	Train/Validation/Test

Figure 1: Training configuration and hyperparameter selection for YOLOv8 pedestrian detection model. The systematic search across 25 configurations explored learning rates from 0.0001 to 0.01 and training epochs from 10 to 50. The selected configuration (LR=0.0005, 50 epochs) achieved optimal validation performance while maintaining computational efficiency suitable for edge deployment. Fixed parameters were chosen based on hardware constraints and YOLOv8 architectural requirements.

The hyperparameter search was conducted systematically across 25 training configurations, with each run logged and validated using standard detection metrics (mAP50, precision, recall). The selected configuration (LR=0.0005, 50 epochs) maximized validation mAP50 while maintaining stable precision/recall characteristics suitable for safety-critical deployment.

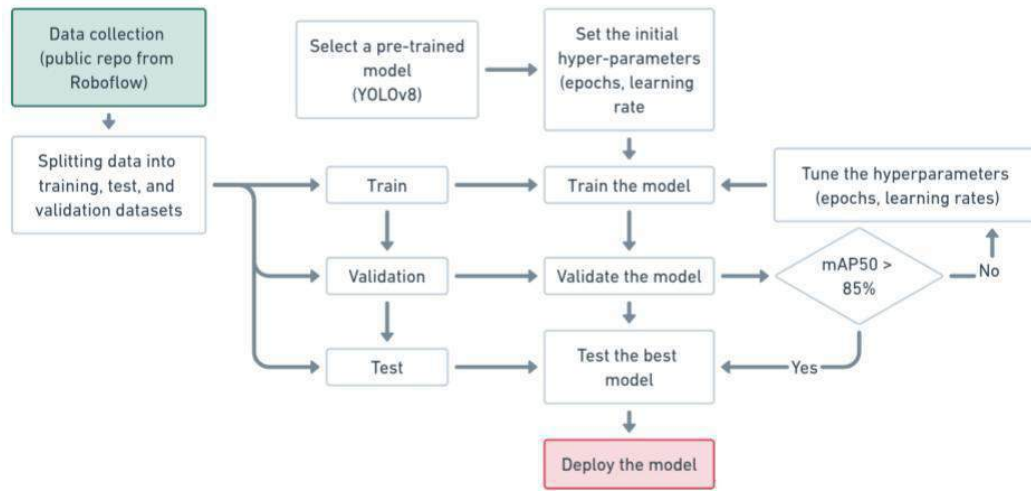


Figure 2: Workflow diagram for training and deploying the YOLOv8 deep learning model used for pedestrian detection. The process begins with data collection and splitting, followed by model selection and hyperparameter tuning. The model is iteratively trained and validated until the desired accuracy threshold ($mAP50 > 85\%$) is achieved. Once validated, the best-performing model is tested and deployed.

Device

To deploy and test the model for a real-world use case, a prototype of a device was built that can be installed on street intersections. To build the device, Raspberry Pi OS was installed on a computer and placed onto a microSD card²⁰. The card was then inserted into the Pi, and the Pi was connected to the servo board using the pins on the Pi. The control code was uploaded to the Pi, the pan-tilt system was assembled, and the motor cables were connected to the servo board²¹.

An acrylic sheet was cut to mount the pan-tilt system, camera, and LEDs. The LEDs were wired to the relays, and the relays were connected to the servo board and a 9V battery. The camera ribbon cable was connected to the Pi, the Pi and servo board were mounted into the case, and wires were routed for organizational purposes. An aluminum T-structure was built for support, and the Pi and pan-tilt were mounted on opposite sides. A 5kg weight was screwed onto the base for balance. The relays were mounted with double-sided tape and Velcro, the different batteries were attached with Velcro for ease of removal, and the cables were organized with enough slack for movement.

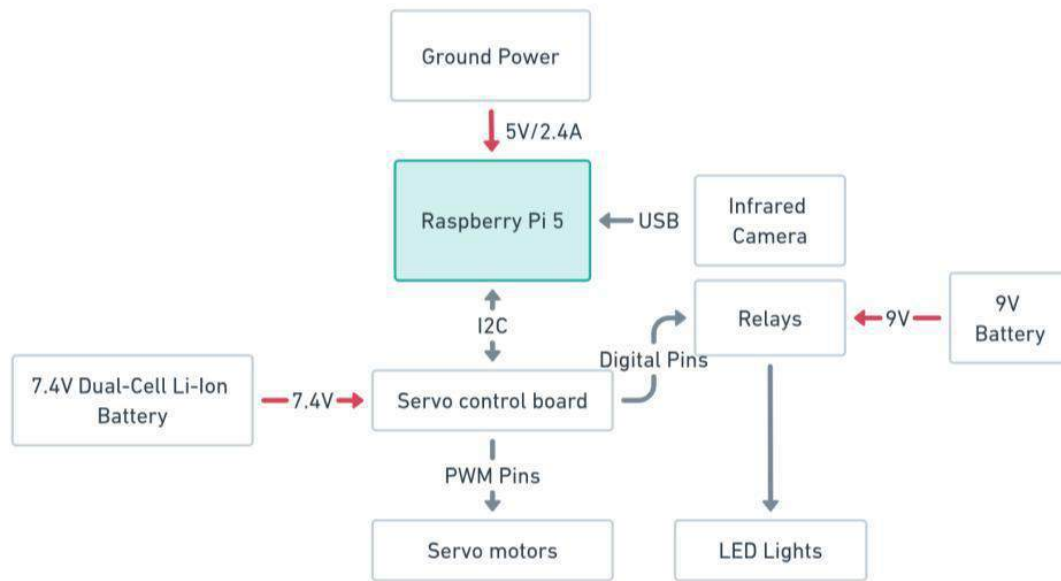


Figure 3: Wiring diagram for the AI Night Vision Crosswalk Guardian device, illustrating the hardware integration between the Raspberry Pi 5, night vision camera, LED lighting system, servo motors, control board, and power sources. Power distribution is managed through separate 7.4V and 9V batteries, with the Raspberry Pi drawing 5V/2.4A from ground power. Control signals are passed through I2C and digital pins to activate relays and servo mechanisms in response to detection events.

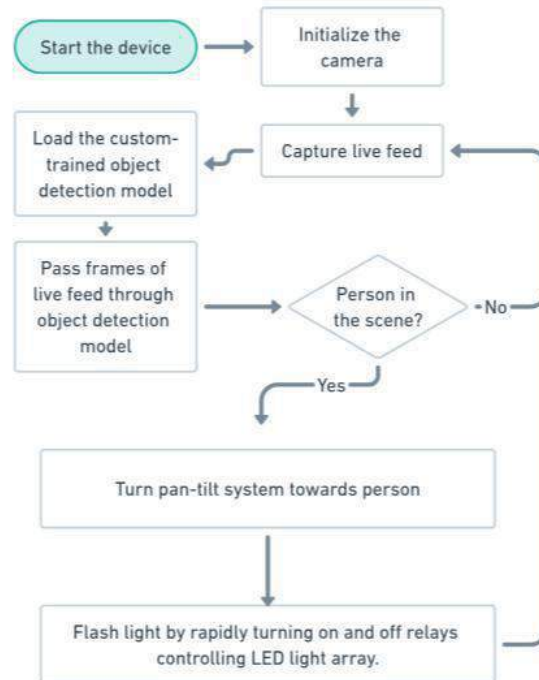


Figure 4: Software control flow for the AI Night Vision Crosswalk Guardian system. Upon device startup, the camera is initialized and begins capturing a live video feed. Each frame is passed through a custom-trained YOLOv8 object detection model developed in this research. If a person is detected, the system activates the pan-tilt servo mechanism to direct the light toward the subject and triggers a flashing LED sequence via relay control. This real-time loop enables responsive pedestrian illumination for enhanced nighttime safety.

Results

Validation results

To determine the most effective model configuration, we ran 25 experiments using YOLOv8 with varying learning rates and training epochs. As visualized in the validation performance graphs (Figures 5 and 6), models validated with a learning rate of 0.0005 consistently showed the best accuracy. The final selected model was validated for 50 epochs with a learning rate of 0.0005, achieving a mean Average Precision (mAP50) of 89.9% on the validation set. Training beyond 50 epochs did not significantly improve performance and risked overfitting.

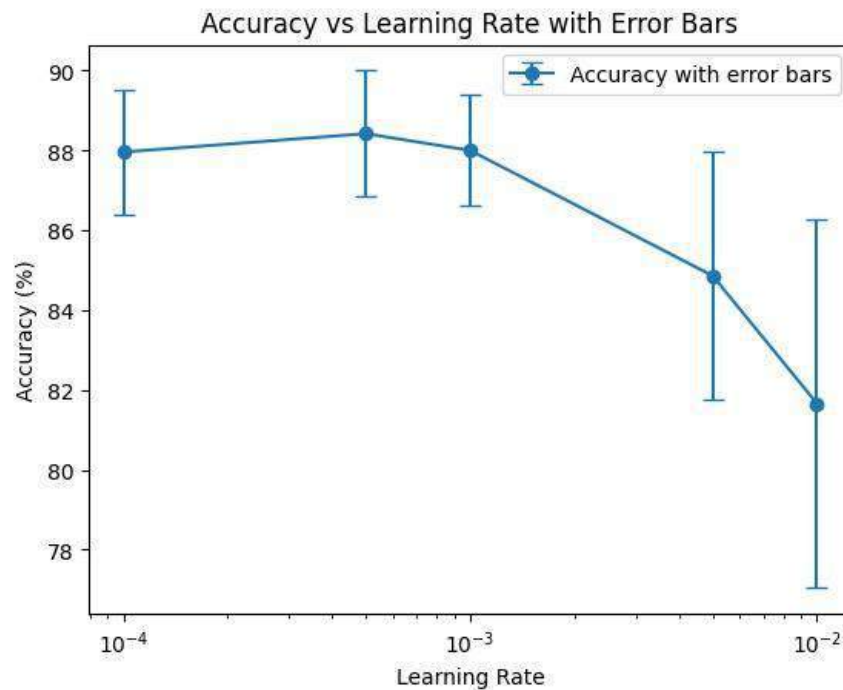


Figure 5: Accuracy vs learning rate with error bars. Error bars show the variation of accuracy for various epochs with the learning rate.

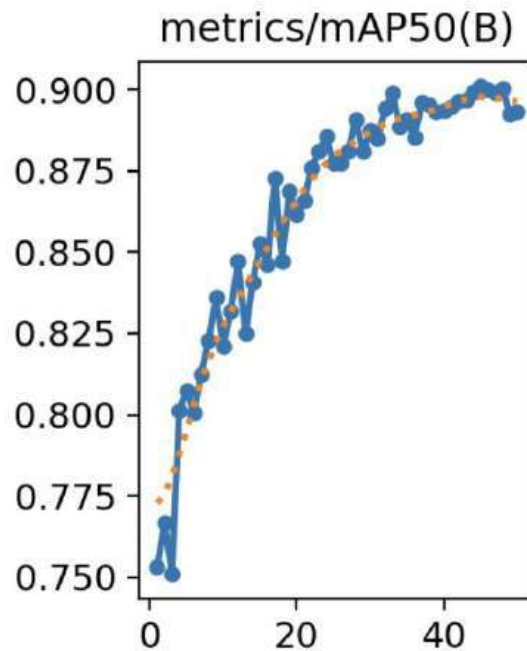


Figure 6: Training performance curve showing mAP50 improvement across epochs for the YOLOv8 model. The plot illustrates a consistent upward trend in mean Average Precision (mAP50), beginning at ~ 0.75 and peaking above 0.90 by epoch 45. The curve flattens after epoch 40, indicating diminishing returns and suggesting an optimal training range between 40–50 epochs to balance learning performance and overfitting risk.

The graph above is a key visualization used to evaluate the object detection model's learning performance over time. It plots the mean Average Precision (mAP) at a 50% Intersection over Union (IoU) threshold on the y-axis against the number of training epochs on the x-axis. This metric reflects how well the model detects and classifies objects as training progresses. A clear upward trend in the graph indicates that the model becomes increasingly accurate with additional training. In the case of this project, the model was trained for up to 50 epochs with a learning rate of 0.0005, and the results show consistent improvement in mAP50 scores as the epoch increases.

However, the performance gains began to plateau around 40 epochs, suggesting that the model had reached a point of diminishing returns. Beyond this point, further training produced minimal improvements and introduced the potential risk of overfitting—where the model memorizes training data rather than generalizing well to new examples. Thus, training between 40 to 50 epochs appears to be optimal for balancing learning depth and generalization. This graph serves as a critical tool for assessing the impact of training parameters and guiding model optimization decisions.

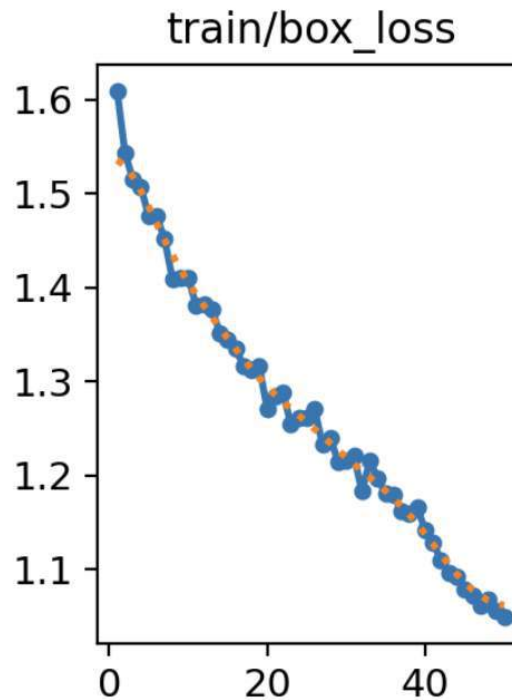


Figure 7: Box loss (accuracy of predicted bounding boxes) curve over 50 training epochs for the YOLOv8 object detection model. The graph shows a steady decline in box loss from approximately 1.6 to below 1.1, indicating progressive improvement in the model's ability to localize objects accurately. The smooth downward trend suggests effective learning without signs of overfitting during training.

The box loss graph visually represents the progression of the model's performance over time, with the x-axis showing the epoch numbers and the y-axis showing the corresponding values of the box loss. This graph serves as a quantitative measure of the error in the predicted bounding box coordinates, capturing how the model's ability to accurately localize objects evolves during the training process. As the model iterates through each epoch, the box loss indicates the degree to which the predicted bounding boxes deviate from the ground truth, with the goal being a reduction in this loss as the model learns to make more precise predictions.

The model trained with a 0.0005 learning rate and 50 epochs had the best detection accuracy (89.9%). It reached an optimal balance between learning and overfitting. An mAP50 score of 89.9% is very high, but the best performance was at LR = 0.0005 and not at 0.0001. Generally, lower LR delivers better accuracy, as it allows the model to converge more gradually and avoid overshooting optimal weights. However, in this case, the model did not strictly follow this trend. This means LR=0.0001 may have slowed convergence too much or caused the model to be suboptimal. This highlights the importance of hyperparameter tuning rather than relying solely on conventional expectations.

Test Results

The selected model was then evaluated on the test dataset. Using Ultralytics 8.3.152, the best model (best.pt) achieved a precision value of 0.956 and a recall value of 0.717. The model achieved an mAP50 of 0.846 and an mAP50–95 of 0.595. The model's inference speed was about 430 ms/image on the CPU, an AMD EPYC 7B12 without a discrete GPU. These results

confirmed the model's reliability for real-time detection, even on low-power edge devices like the Raspberry Pi.

Device Performance Tests

To evaluate the effectiveness of the deep-learning Night Vision Crosswalk Guardian, the final trained model (best.pt) was first deployed on the device prototype and tested in a low-light indoor environment simulating nighttime street conditions. A total of 10 preliminary trials were conducted in this controlled setting to establish baseline functionality and ensure the system operated as intended.

Building on the successful indoor results, we then expanded testing to real-world conditions. Forty additional trials were conducted outdoors on a public street at night, under naturally low-light environments. Participants ranged in height from approximately 150 cm to 190 cm, wore both light and dark clothing, and varied in body size. They approached either on foot or by bicycle, with some trials involving two to four participants in the frame at once to evaluate multi-subject detection. In every case, the LED array activated immediately upon a participant entering the camera's wide field of view and remained on until they exited, while consistently ignoring nearby passing vehicles. Testing was performed under mild outdoor weather conditions, with environmental noise such as headlights and street activity present. These outdoor trials demonstrated stable detection performance across all approach angles and confirmed the infrared camera's effectiveness in realistic, uncontrolled environments. The device still reliably detected subjects even while there was foliage blocking the view of the device. Future work would extend the testing of the device's performance under varying weather conditions, such as rain, wind, and fog, to further validate its robustness.

Model vs Real-World Performance Explanation

An important distinction exists between the model's frame-level recall (0.717 on test dataset) and the observed event-level performance (100% detection in field trials where power supply was adequate). Frame-level recall measures detection success on individual images, while event-level performance reflects the system's ability to activate illumination during crossing events.

In our 50 field trials (10 indoor + 40 outdoor), we observed successful activation in all events where power supply was adequate, yielding an event-level recall of 1.0 under tested conditions. Using conservative statistical bounds (rule-of-three), this suggests an event-level recall >94% with 95% confidence for similar mild weather, controlled scenarios. The higher event-level performance likely results from temporal redundancy—multiple frames capture each crossing event, requiring only one successful detection to trigger illumination.

Summary of the above section

The combination of a high-performing YOLOv8 model and a responsive physical system validates the effectiveness of the deep-learning Night Vision Crosswalk Guardian. The model not only maintained high accuracy in software testing but also demonstrated reliable real-world

behavior under hardware constraints, reinforcing its potential for scalable deployment in smart infrastructure.

Discussion

Technical Performance Analysis

The YOLOv8 experiments demonstrated strong performance across both software and hardware environments. Initial trials used epochs ranging from 10–50 and learning rates between 0.0001–0.01. Results showed that mAP50 values plateaued beyond 40 epochs at nearly all learning rates, so training was capped at 50 epochs. The model achieved high precision (0.956) and strong mAP50 (0.846) on the test set, confirming effective training and generalization. Although recall (0.717) was slightly lower, this trade-off was acceptable to minimize false positives in real-time, safety-critical contexts.

In the initial round of hardware testing, the system reached 80% detection accuracy in low-light conditions; failures were traced to wiring and power issues rather than model flaws. Once corrected, the system consistently detected pedestrians immediately upon entering the frame, within the real-time alerting threshold. These findings highlight three key insights: First, the YOLOv8 model architecture is highly effective for embedded applications such as the Night Vision Crosswalk Guardian, even on non-GPU hardware; Second, hardware reliability is as critical as model accuracy, requiring careful attention to build quality and power management; and third, the integration of deep-learning detection with responsive physical infrastructure demonstrates that AI-powered crosswalk safety systems can be both accurate and practical for real-world deployment.

Privacy-Preserving Analytics and Deployment Considerations

The system supports optional privacy-preserving situational analytics through on-device inference without facial recognition or persistent video storage. Aggregate pedestrian counts (e.g., crossings per hour) can be logged locally for traffic planning purposes while maintaining individual privacy through anonymous detection-only processing. Unlike traditional surveillance systems, this approach prioritizes safety intervention over monitoring, ensuring that individual privacy remains protected while still providing valuable urban planning insights.

Many urban crosswalks lack adjacent freestanding poles or suitable mounting infrastructure. For such locations, we designed three installation approaches: (1) dedicated 4-6 meter poles with underground electrical connections, (2) building-mounted brackets with crosswalk-directed orientation, and (3) traffic light integration using existing signal infrastructure. Each approach maintains the required 15-foot minimum mounting height and includes a calibration protocol to ensure the camera's field-of-view correctly covers the crosswalk polygon region-of-interest.

System	Type	Key Tech	Cost	Notes
Night Vision Crosswalk Guardian (This Work)	Prototype	IR + YOLOv8 + LED	~\$700	Real-time detection + illumination
Park et al. (2024)	Academic	RGB+IR fusion	N/A	High accuracy, vehicle-focused
Ghose et al. (2019)	Academic	Thermal + saliency	N/A	Boosts low-light detection
Nowosielski et al. (2020)	Academic	Low-cost IR DNN	Low	Portable, less accurate
FLIR Thermal Cameras	Commercial	Thermal imaging	\$2,000+	Proven, expensive
Lightguard IRWL	Commercial	In-road LEDs	\$1,500–3,000	Strong alerts, no AI
Smart Bollards	Commercial	LED warnings	\$1,200–2,500	Simple, no detection

Figure 8: Comparative analysis of the Night Vision Crosswalk Guardian prototype against academic low-light pedestrian detection systems and commercial crosswalk safety solutions, highlighting key technologies, costs, and functional capabilities

Cost-Benefit and Feasibility Analysis

While solar-powered, motion-activated streetlights are widely available and relatively inexpensive, they are designed to trigger illumination for any movement (including vehicles) and to provide diffuse area lighting rather than precise, targeted illumination. In contrast, the Night Vision Crosswalk Guardian responds exclusively to pedestrians and cyclists and directs a concentrated light beam at pedestrians or cyclists during a crossing event, ensuring that approaching drivers are immediately alerted to their presence. This targeted approach reduces wasted energy and enhances visual saliency compared to background illumination for all motion. The continuously running part of the device consumes a maximum of 10 watts, and since the floodlight will turn on only when needed, it will result in power being saved.

From a cost perspective, a final, polished version of this prototype, with a custom-machined enclosure, a high-end NVIDIA Jetson Nano edge computer, and a high-power 100W floodlight, estimates the final unit price at approximately \$700. This is comparable to many mid- to high-end CCTV systems currently deployed for traffic surveillance, which typically range from \$800–\$2000. Unlike traditional CCTV cameras, the Guardian integrates both detection and active illumination in one device, offering dual-use functionality as both a crosswalk safety system and a privacy-preserving analytics platform.

Importantly, our cost analysis assumes up to one unit per individual crosswalk, recognizing that complex intersections with more than 2 pedestrian crossing paths may require multiple devices for complete coverage. The \$700 per-unit estimate remains competitive when compared to multi-point IRWL installations that typically require 4-8 embedded lights per crossing and smart bollards that often require multiple installations per crosswalk, substantially increasing total system cost. The device should be mounted at least 15 feet high and has a very large detection and illumination range due to the high-quality camera and floodlight. The expected operational lifespan is 5–10 years before requiring replacement of the floodlight component.

When considering performance, coverage, and lifetime maintenance, the Guardian

represents a competitive and scalable solution for municipalities seeking to enhance nighttime pedestrian safety. The final device will be weatherproof with a waterproof enclosure, with all connections bridged with simple Anderson connectors to ensure durability and repairability. The device will be powered through ground power to ensure a stable supply of electricity, with an option for solar power available if needed during power outages.

Response Time and Driver Perception

The system's detection speed allows for instant driver visibility. In indoor testing vs real-world testing, the device went from detecting in under 3 seconds to almost instantly, due to the installation of a camera with a faster connection. To further reduce latency, future iterations will employ faster edge processors (e.g., NVIDIA Jetson Orin Nano) and optimized inference pipelines. Complementary measures such as vehicle-to-infrastructure (V2I) communication or auxiliary driver alerts could also be integrated to extend safety coverage.

Future Work

Environmental Testing Requirements

Current field trials were conducted under mild weather conditions (clear skies, minimal wind). The system's performance under adverse conditions including fog, rain, and wind requires systematic evaluation. Future work will implement dataset augmentation with weather-degraded imagery, controlled winter testing across precipitation and visibility conditions, and lens treatments and enclosure modifications to maintain optical clarity. These environmental robustness studies will establish operational boundaries and inform deployment guidelines for different climatic regions.

Future research will also focus on the development of a multi-sensor fusion model that combines RGB and thermal imaging for enhanced perception in varying lighting conditions.

Finally, it will go under large-scale municipal deployment studies to validate effectiveness across diverse urban geometries and traffic patterns.

Conclusion

This research demonstrates the feasibility of an AI-driven, event-aware pedestrian detection system that enhances crosswalk safety through targeted illumination. Unlike generic motion-activated lighting, our class-gated approach (person-only detection) combined with crosswalk-specific deployment geometry provides precise, privacy-preserving safety intervention while minimizing false activations and energy waste.

The YOLOv8 deep-learning model achieved an impressive 89.9% detection accuracy (mAP50) on the validation dataset. In uncontrolled low-light testing, the prototype successfully detected pedestrians and activated lights in 100% of experiments where power supply was adequate, detecting as soon as the subject entered the frame.

While AI-based human detection has been extensively explored in security and surveillance, there is limited research on its application for pedestrian and cyclist visibility in

road safety. This project presents a novel integration of deep-learning–based object detection with real-world safety infrastructure, resulting in an AI-powered floodlight system running on an edge computing device, specifically designed to improve visibility at crosswalks.

Unlike traditional surveillance-based models, this project does not rely on facial recognition or personal identification, ensuring privacy protection while improving safety. Instead, it detects the presence of pedestrians and cyclists, making roads safer without compromising individual privacy.

The system represents an improvement rather than a complete solution to pedestrian safety challenges. While targeted illumination addresses visibility gaps during crossing events, additional risk factors including driver behavior, road design, and environmental conditions require complementary interventions for comprehensive safety enhancement.

Works Cited

1. Bicycle Safety <https://www.nhtsa.gov/book/countermeasures-that-work/bicycle-safety>.
2. FHWA Lighting Handbook, https://highways.fhwa.dot.gov/sites/fhwa.dot.gov/files/2023-05/FHWA-Lighting-Handbook_0.pdf 2023.
3. Pedestrian Lighting Primer, https://highways.fhwa.dot.gov/sites/fhwa.dot.gov/files/2022-09/Pedestrian_Lighting_Primer_Final.pdf April 2022.
4. Reducing Late-Night/Early-Morning Intersection Crashes By Providing Lighting, https://highways.fhwa.dot.gov/sites/fhwa.dot.gov/files/dot_49237_DS1.pdf Dec 2009.
5. Teledyne Flir, https://www.flir.com/?srsltid=AfmBOopMobSY-wZd0PbIzQv73sKmhpfPTnJBMiXAZT523afllu_SaDmo.
6. Lightguard Systems Smart Crosswalk IWRL, Lightguard Systems, <https://www.lightguardsystems.com/smart-crosswalk-in-roadway>.
7. Crosswalk Signals Smart Pedestrian Warning Bollards, Crosswalk Signals, <https://www.trafficolution.cn/product/smart-pedestrian-warning-bollards>.
8. J. Park, B. K. Thota, K. Somashekar. Sensor-Fused Nighttime System for Enhanced Pedestrian Detection in ADAS and Autonomous Vehicles. *Sensors* 2024, 24(14), 4755, <https://www.mdpi.com/1424-8220/24/14/4755>.
9. D. Ghose, S. M. Desai, S. Bhattacharya, D. Chakraborty, M. Fiterau, T. Rahman. Pedestrian Detection in Thermal Images using Saliency Maps. *arXiv:1904.06859v1*, <https://arxiv.org/abs/1904.06859>.
10. Q. Liu, Z. He, X. Li, Y. Zheng. PTB-TIR: A Thermal Infrared Pedestrian Tracking Benchmark. *IEEE Transactions on Multimedia*, vol. 22, no. 3, pp. 666-675, March 2020, <https://ieeexplore.ieee.org/document/8798885>.
11. H. Kulhandjian, J. Barron, M. Tamiyasu, M. Thompson, M. Kulhandjian. Pedestrian Detection and Avoidance at Night Using Multiple Sensors and Machine Learning. 2023 International Conference on Computing, Networking and Communications (ICNC), Honolulu, HI, USA, 2023, pp. 165-169, <https://ieeexplore.ieee.org/document/10074081>.
12. S. Geng. Infrared image pedestrian target detection based on Yolov3 and migration learning. *arXiv:2012.11185v1*, <https://arxiv.org/abs/2012.11185>.
13. T. Akilan, H. Vachhani. Low-light Pedestrian Detection in Visible and Infrared Image Feeds: Issues and Challenges. *arXiv:2311.08557v3*, <https://arxiv.org/abs/2311.08557>.
14. A. Nowosielski, K. Małeck, P. Forczmański, A. Smoliński, K. Krzywicki. Embedded Night-Vision System for Pedestrian Detection. *IEEE Sensors Journal*, vol. 20, no. 16, pp. 9293-9304, 15 Aug. 15, 2020, <https://ieeexplore.ieee.org/document/9062586>.
15. A. Raza, S. A. Chelloug, M. H. Alatiyyah, A. Jalal, J. Park. Multiple Pedestrian Detection and Tracking in Night Vision Surveillance Systems. *CMC*, vol. 75, no. 2, pg. 3275–3289, 2023, <https://www.techscience.com/cmc/v75n2/52020/pdf>.

- 16.X. Wang, J. Chen, Z. Wang, W. Liu, S. Satoh, C. Liang, C. Lin. When Pedestrian Detection Meets Nighttime Surveillance: A New Benchmark. Twenty-Ninth International Joint Conference on Artificial Intelligence Main track, pg. 509-515, 2020, <https://www.ijcai.org/proceedings/2020/71>.
- 17.Night Vision Computer Vision Model. sagemcom, <https://universe.roboflow.com/sagemcom-nh1r3/night-vision-mzy8e>.
- 18.Ultralytics YOLOv8. Multiple authors. <https://docs.ultralytics.com/models/yolov8/>.
- 19.P. Skalski. How to Train YOLOv8 Object Detection on a Custom Dataset. Roboflow Blog, Jan 10. 2023, <https://blog.roboflow.com/how-to-train-yolov8-on-a-custom-dataset/>.
- 20.Raspberry Pi Documentation. Multiple authors. <https://www.raspberrypi.com/documentation>.
- 21.SunFounder Robot HAT. Multiple authors. <https://docs.sunfounder.com/projects/robot-hat-v4/en/latest/>.

Nanotechnology In Metastatic Breast Cancer By Darara Meskela

Acknowledgements

Thank you for the guidance of Dr. Jon Ostolaza Paraiso from the University of Cambridge in the development of this research paper.

Abstract

Breast cancer is still one of the most prevalent and lethal forms of cancer that affect millions of women worldwide, with emerging cases predicted in the coming decades. While traditional treatment methods such as surgery, radiation, and chemotherapy have all taken pivotal roles in treating the disease, they are often associated with significant limitations—such as systemic toxicity, limited selectivity for neoplastic tissue, and emerging multidrug resistance. Here, nanotechnology has emerged as a promising field with new solutions through nanoparticle-mediated drug delivery systems to enhance therapeutic specificity and reduce side effects.

This paper discusses the ways in which the various nanoparticles like liposomes, polymeric nanoparticles, gold nanoparticles, carbon nanotubes, and quantum dots have revolutionized the area of breast cancer therapy. These nanoscale platforms operate by mechanisms that encompass passive targeting by the EPR effect, active targeting by ligand-receptor interactions, and stimuli-responsive release by tumor-specific stimuli like pH or temperature. These mechanisms, in combination, enhance drug bioavailability considerably, reduce toxicity to normal tissues, and enhance drug retention in cancer cells. Furthermore, theranostic application of nanoparticles—combining diagnostic imaging and therapeutic drug delivery on a single platform—has enabled more individualized and adaptive treatment strategies based on personal patient profiles.

However, challenges that include long-term safety, biodegradability, scale-up manufacturability, clinical efficacy validation, and regulatory pathways linger to hinder comprehensive clinical translation. This article does not only discuss the scientific concept of nanomedicine but also critically examines its current limitations and future directions. Lastly, continued research and multidisciplinary investigation hold the solution to unlocking the full potential of nanotechnology-based intervention in achieving increased survival and quality of life for breast cancer patients worldwide.

[I] Introduction to Nanotechnology in Breast Cancer

Breast cancer remains one of the most frequent and lethal types of cancer across the world, affecting millions of individuals annually and posing a burden to healthcare systems in both industrialized and less developed nations. According to the International Agency for Research on Cancer (IARC), over 2.3 million new breast cancer diagnoses took place worldwide in 2022 and resulted in an estimated 670,000 deaths. If current trends persist, projections suggest that in 2050, the incidence of breast cancer may increase to over 3.2 million cases per year, and the mortality burden to 1.1 million. This projected rise in incidence and mortality is attributable to a variety of factors, including population aging, environmental exposures, changes in diet and

lifestyle, obesity, alcohol consumption, late childbearing, and limited access to early detection or high-quality treatment, particularly in low- and middle-income countries.

Breast cancer is a heterogeneous disease, with distinct molecular subtypes—e.g., hormone receptor-positive (HR+), human epidermal growth factor receptor 2-positive (HER2+), and triple-negative breast cancer (TNBC)—each with varying prognoses, treatment responses, and biology. While current treatments such as surgery, radiation, endocrine therapy, and chemotherapy have dramatically improved the outcomes of breast cancer patients during the past few decades, they are frequently accompanied by considerable drawbacks. These are largely systemic toxicity, multidrug resistance, non tumor specificity, and toxic side effects that are harmful to the patient's quality of life.

Traditional chemotherapeutic agents, for instance, attack proliferating cells indiscriminately irrespective of their organ of origin and produce off-target cytotoxicity in normal tissues like bone marrow, gastrointestinal tract, and hair follicles. Furthermore, the indiscriminate character of these therapies typically requires high doses to be systemically administered to achieve therapeutic levels within the tumor bed, increasing toxicity without necessarily ensuring efficacy. Such limitations have led to an international scientific movement aimed at the development of more precise, tumor-directed delivery systems—the dawn of nanomedicine.

Nanotechnology presents a new and versatile platform to revolutionize breast cancer therapy through the use of engineered nanocarriers to deliver chemotherapeutic agents, gene therapies, or biologics to tumor cells. Nanomaterials, typically in the range of 1-100 nanometers in diameter, are engineered to cross biological barriers, extend circulation times, and target tumor tissue selectively through the enhanced permeability and retention effect or through active targeting of cancer-specific receptors.

Most notably, nanotechnology allows for functional tailoring of therapeutic payloads. Surface engineering of nanoparticles with targeting ligands (such as monoclonal antibodies, peptides, or aptamers) enables researchers to increase cellular uptake by tumor cells while minimizing collateral damage to normal tissue. Additionally, nanocarriers may be programmed to respond to internal tumor stimuli—pH, redox gradients, or enzymatic activity—or external stimuli like heat or magnetic fields, allowing for controlled and on-demand drug release.

In addition to drug delivery, nanotechnology also plays an important role in cancer diagnosis and monitoring. Theranostic nanoparticles that integrate diagnostic imaging and therapeutic capabilities have enabled real-time monitoring of tumor growth, drug distribution, and treatment response—enabling highly individualized medicine.

With growing sophistication in its resistance and intricacy, the use of nanotechnology in clinical oncology represents a paradigm change in our management and treatment modalities. Some nanoparticle-formulated drugs, such as Doxil and Abraxane, are currently FDA approved. Many others continue to be researched clinically and preclinically and reflect both promise and pitfalls of the field. These include issues of long-term toxicity, biodistribution, biodegradability, cost efficacy, and regulation.

This article is designed to explore the revolutionizing capabilities of nanotechnology in breast cancer therapy through the focus on the nature of nanoparticles employed, the delivery systems, clinical benefits, and the multi-aspectual challenges required to be met to translate these novel technologies from the bench to the bedside.

II. Types of Nanoparticles Used in Breast Cancer Treatment

Liposomes

Liposomes are spherical vesicles with one or more lipid bilayers, quite analogous to the cellular membrane structure. Because of this unique biomimetic attribute, liposomes can encapsulate hydrophobic as well as hydrophilic drugs efficiently, enhancing their solubility, pharmacokinetics, and overall bioavailability in the body. Liposomes have drawn considerable attention for their potential to reduce systemic toxicity and allow selective drug delivery, most prominently in the context of cancer therapy. Perhaps the most prominent example is Doxil, a liposomal formulation of the anthracycline chemotherapeutic agent doxorubicin. The product has shown considerable clinical effectiveness in breast cancer treatment, most notably due to its reduced cardiotoxic side effects compared to its free drug counterpart.

Doxil's efficacy is based, in great part, on its ability to capitalize on the enhanced permeability and retention (EPR) effect, an inherent phenomenon in solid tumors (Barenholz, 2012). The EPR effect is composed of two predominant mechanisms:

Increased permeability: The tumor vasculature is typically disordered and irregular with fenestrae between 1 and 100 nanometers. Leaky vessels facilitate passive diffusion of nanoparticles, i.e., liposomes, through the tumor interstitium. Contrary to conventional chemotherapy, where medication diffuses non-selectively, treating cancer cells and normal cells alike.

Insufficient lymphatic drainage: The tumor tissue generally lacks or impaired lymphatic system, which debilitates removal of nanoparticles from tumor microenvironment. As a result, liposomes remain within the tumor location for extended time and ensure extended duration and site-targeted drug delivery.

In order to additionally increase circulation time and reduce detection by the immune system, liposomes are conventionally PEGylated—coated with polyethylene glycol—such that they can escape opsonization and phagocytosis. The additive effect of targeted delivery, enhanced tumor deposition, and increased systemic circulation makes liposomes an unbeatable tool in breast cancer therapy. Beyond doxorubicin, liposomal delivery is now being explored for other chemotherapeutic agents, gene therapy vehicles, and even RNA-based drugs, and new avenues are opening to precision medicine.

Polymeric Nanoparticles

Polymeric nanoparticles, composed of biodegradable polymers such as poly(lactic-co-glycolic acid) (PLGA), chitosan, and polylactic acid (PLA), offer a highly advanced platform for controlled delivery of anticancer agents. These particles are usually

engineered to degrade in a predictable manner in the body, releasing their payload over a prolonged period. This controlled slow and continuous release process is especially beneficial for drugs with narrow therapeutic windows, like paclitaxel, which is utilized widely in breast cancer treatment (Danhier et al., 2012).

PLGA nanoparticle-encapsulated drug is protected against enzymatic breakdown in the circulation prior to premature degradation, leading to drug stabilization until the arriving nanoparticle at the tumor site. In addition, the nanosurface is functionalized using targeting moieties—e.g., folic acid, antibody, or peptide—to enhance interaction with cancer cell receptors. pH- and temperature-sensitive linkages also provide the nanoparticle stimuli-responsive functionality allowing the release of the cargo to match the condition of the target tumor microenvironment.

The applications of polymeric nanoparticles extend beyond chemotherapy. They are being explored for co-delivery systems, where two or more therapeutic agents—such as a chemotherapeutic and a gene silencer—are delivered simultaneously to achieve maximum therapeutic effects. Their biocompatibility, tunability, and ability to circumvent multidrug resistance mechanisms make them a prime candidate for next-generation oncologic treatments.

Gold Nanoparticles

Gold nanoparticles (AuNPs) are a versatile family of nanomaterials with excellent potential in the treatment as well as the diagnosis of breast cancer. Due to their surface plasmon resonance properties, AuNPs can absorb and scatter light efficiently, an effect that is capable of being harnessed in photothermal therapy (PTT). In this approach, AuNPs accumulate in the tumor, and when exposed to near-infrared (NIR) light, they convert the absorbed light into localized heat that selectively kills cancerous tissues (Huang et al., 2011).

Apart from their photothermal nature, gold nanoparticles are very good carriers for a wide variety of therapeutic and diagnostic agents. Their surfaces can easily be functionalized with low-molecular-weight molecules, antibodies, nucleic acids, and drugs, which renders them very customizable. In the case of breast cancer, AuNPs tend to be functionalized with monoclonal antibodies against overexpressed receptors such as HER2, allowing for active targeting and increased selectivity.

The surface chemistry, size, and shape of AuNPs dictate their biodistribution, cellular uptake, and clearance from the body. Small particles tend to have longer circulation times and improved tissue penetration, whereas larger particles can offer increased drug loading capacity. This enables researchers to tailor gold nanoparticles for both systemic and local treatments. Besides, their inertness, low toxicity, and compatibility with imaging agents such as computed tomography (CT) and magnetic resonance imaging (MRI) have made them the core tool in cancer theranostics.

Carbon Nanotubes and Quantum Dots

Carbon nanotubes (CNTs) are cylindrical nanostructures composed of rolled graphene sheets, prized for their extraordinary strength, electrical conductivity, and thermal stability. Their rod-like morphology makes it relatively easy for them to penetrate cellular membranes, facilitating the intracellular delivery of therapeutic compounds, including anticancer drugs and nucleic acids (Liu et al., 2008). Functionalization with hydrophilic polymers or targeting ligands increases their solubility and biocompatibility, making it easier for them to transport more efficiently within biological systems. In cancer of the breast, CNTs have been studied as drug carriers for doxorubicin, paclitaxel, and siRNA, showing their application in both chemotherapy and gene silencing therapy.

Quantum dots (QDs) are semiconductor nanocrystals with outstanding optical properties. They emit bright, stable fluorescence when excited and can be engineered to emit light over a wide range of wavelengths. These characteristics make QDs extremely useful in imaging applications such as tracking tumor growth, sentinel lymph node detection, and guiding surgical resections (Gao et al., 2005). Moreover, when combined with tumor-targeting ligands, QDs can bind specifically to cancer cells, enabling high-resolution and real-time visualization of tumor margins.

While they are promising, both QDs and CNTs have major issues with long-term biocompatibility and toxicity. Carbon nanotubes can induce oxidative stress, inflammation, and granuloma when not cleared by the body in a proper manner. Similarly, quantum dots are often made up of heavy metals such as cadmium, which can pose the risk of systemic toxicity. Hence, the path to clinical translation is through extensive toxicological evaluation and optimization of these nanocarriers for safety without losing efficacy.

Briefly, the field of nanoparticles for the treatment of breast cancer is wide and growing. Each category—liposomes, polymeric nanoparticles, gold nanoparticles, carbon nanotubes, and quantum dots—offers something unique, and research is under way to offset each of their limitations. Taken together, they represent an exciting array of tools for enhancing accuracy, efficiency, and safety in breast cancer therapy.

III. Mechanisms of Nanoparticle Drug Delivery

Active Targeting

Active targeting is a sophisticated method in nanoparticle-based drug delivery systems, where nanoparticles are engineered to target and bind specifically to cancer cells by capitalizing on overexpressed specific molecular markers on their surface. Unlike passive targeting, which relies on the physiological heterogeneity between normal tissue and tumor tissue, active targeting enhances specificity by utilizing ligands, antibodies, or peptides conjugated on the nanoparticle surface. These targeting agents will bind to the overexpressed receptors on the cancer cells, delivering a higher dose of the therapeutic at the disease site and less toxicity to normal tissues.

Antibody-based targeting represents one of the most important instances of active targeting. For this approach, monoclonal antibodies are used to functionalize nanoparticles to target certain

tumor-specific antigens. One of the most widely studied targets is the human epidermal growth factor receptor 2 (HER2) overexpressed in approximately 15-20% of breast cancers. Nanoparticles with Herceptin (trastuzumab), a monoclonal antibody against HER2, labeling have been reported to significantly improve delivery and action of chemotherapeutic agents such as doxorubicin or paclitaxel in HER2-positive breast cancer. This method not only enhances cellular uptake of drug-loaded nanoparticles but also leads to apoptosis and inhibits tumor growth through suppression of HER2-signaling pathways.

Another active targeting method is peptide-based targeting. Such peptides as RGD (arginine-glycine-aspartic acid) are highly specific to bind with integrins, particularly $\alpha v \beta 3$ integrins that are overexpressed on tumor-related angiogenic blood vessels. RGD-functionalized nanoparticles were reported to effectively target the tumor vasculature, allowing enhanced tissue penetration and drug delivery within the tumor microenvironment. Furthermore, the minimal size and ease of synthesis of peptides render them attractive candidates for nanoparticle modification.

Ligand-mediated targeting also has an important role in active delivery. Folic acid and transferrin are frequently utilized ligands because their receptors are overexpressed in most forms of cancer, including aggressive forms of breast cancer. Targeting of the folate receptor (FR) is especially effective against triple-negative breast cancer (TNBC), which is a subtype with ER, PR, and HER2 lack. Transferrin receptors (TfR) are, however, overexpressed on dividing cells since they require more iron. Folate or transferrin ligand-coated nanoparticles exhibit enhanced binding and internalization in cancer cells, resulting in greater delivery of therapeutic payload, reduced systemic side effects, and the ability to circumvent most drug resistance pathways (Kobayashi et al., 2014).

Collectively, active targeting therapies have superior therapeutic impact and decreased systemic toxicity by using selective accumulation of drug-loaded nanoparticles within the tumor environment. These therapies are at the forefront of individualized oncology medicine with superior potential for enhanced therapeutics and reduced damage.

Stimuli-Responsive Delivery

Stimuli-responsive delivery systems are the latest generation of nanocarriers that can deliver their therapeutic payloads upon exposure to specific external or internal stimuli. These smart delivery systems of drugs can respond to various biochemical stimuli present in the tumor microenvironment, such as pH, temperature, redox potential, enzymatic level, or external triggers like light, magnetic field, or ultrasound. Taking advantage of the unique circumstances present in the tumor, such systems aim further to enhance precision and effectiveness in drug delivery utilizing nanoparticles.

pH-sensitive nanoparticles are one such widely researched variety of internal stimulus-responsive delivery. Solid tumors possess a mildly acidic extracellular environment (pH ~6.5-6.9) compared to normal tissue (pH ~7.4), largely due to altered metabolism and compromised vascular perfusion. pH-responsive nanoparticles are crafted to remain stable at physiological pH but release their drug payload upon exposure to an acidic environment. This

avoids pre-release in the bloodstream and appreciably reduces systemic toxicity. For instance, nanoparticles of lipids or pH-sensitive polymers alter their shape or get hydrolyzed when exposed to acidic pH and trigger release of loaded drugs (Zhang et al., 2019).

Temperature-sensitive nanoparticles offer a further degree of controlled drug delivery. Certain tumors are defined by localized hyperthermia, and nanoparticles that are designed to respond to small temperature increases can release their drug exactly where they are required. Thermosensitive systems are often formulated with materials like poly(N-isopropylacrylamide), which alter solubility or structure at a specific temperature threshold.

In addition, enzyme-responsive nanoparticles exploit the overexpression of specific enzymes in tumor tissues, such as matrix metalloproteinases (MMPs), cathepsins, or esterases. Through the utilization of cleavable linkages in the nanoparticle design responsive to these enzymes, drug release can be precisely controlled at the tumor site. This mechanism adds a further layer of specificity, whereby therapeutic agents are released only where the corresponding enzyme is expressed.

Coupled, stimuli-responsive nanoparticles represent a paradigm shift in precision oncology. Not only do they increase local drug concentration but they provide on-demand release, reduce off-target effects, and enable more reproducible pharmacokinetics. Such platforms are of outstanding promise when combined with passive and active targeting methodologies, leading to multifunctional platforms that integrate treatment with real-time regulation and sensitivity to the tumor environment.

IV. Benefits of Nanotechnology in Breast Cancer Therapy

Enhanced Drug Bioavailability

Another of the principal challenges with conventional breast cancer chemotherapy is that many anticancer drugs are water insoluble, especially hydrophobic drugs such as paclitaxel or docetaxel. Nanoparticles circumvent this challenge by placing such water-insoluble drugs within nanometer-sized vehicles that enhance their aqueous solubility and protect them from premature degradation. This encapsulation not only improves the physicochemical stability of the drug but also delivers it more efficiently to its therapeutic location. Therefore, the overall bioavailability of the drug is significantly improved, allowing a higher proportion of the active ingredient to be deposited in the tumor location where it is needed most (Peer et al., 2007). This is especially required for drugs that must be released over time to be effective while minimizing changes in systemic circulation.

Less Systemic Toxicity

Traditional chemotherapy is typically achieved by systemic delivery of cytotoxic agents, which circulate throughout the body and attack cells in division quickly indiscriminately—hitting tumor cells as well as normal cells in the bone marrow, gastrointestinal tract, and hair follicles. This non-specific delivery often leads to incapacitating side effects such as cardiotoxicity, peripheral neuropathy, and immune suppression. Nanoparticle drug delivery

systems, however, may be formulated to preferentially target tumor tissues via either passive routes such as the enhanced permeability and retention (EPR) effect or active targeting by ligands or antibodies. By localizing the drug payload within the tumor and restricting its delivery to healthy tissues, nanoparticles have a dramatic decrease in the possibility of off-target effects and systemic toxicity (Tiwari et al., 2012). This translates into better patient compliance, reduced disruption of treatment, and thus improved therapeutic response.

Overcoming Multidrug Resistance (MDR)

Multidrug resistance is one of the most formidable challenges in the long-term treatment of breast cancer. It usually occurs when the cancer cells develop mechanisms of pumping out drugs before these drugs have a chance to exert their therapeutic effect, often through overexpression of efflux pumps such as P-glycoprotein (P-gp). Nanotechnology provides a solution to this problem. Nanoparticles can encapsulate chemotherapeutic drugs and shield them from being recognized by such efflux transporters. Further, by facilitating endocytosis-mediated internalization in cancer cells, nanoparticles have the potential to deliver drugs into the intracellular space directly, thus bypassing these resistance mechanisms. Some nanoparticles are also designed to release their load in a controlled manner upon internalization, maximizing drug retention and action in the tumor microenvironment (Chen et al., 2016). This has significant implications for patients with drug-resistant tumors who have exhausted traditional treatment options.

Theranostic Applications

One of the most exciting nanomedicine frontiers is the development of theranostic platforms—multifunctional nanoparticles with both therapeutic and diagnostic capabilities in a single formula. These nanoparticles can be filled with imaging agents such as quantum dots, magnetic nanoparticles, or fluorescent dyes in combination with therapeutic drugs. This can allow real-time monitoring of drug delivery, tumor localization, and therapy response alongside the concomitant delivery of therapy. An integrated strategy like this facilitates personalized medicine through the ability of clinicians to customize treatment regimens according to particular tumor characteristics and responses. Moreover, theranostics reduce the application of invasive interventions, improving patient comfort and clinical workflow efficiencies (Jiang et al., 2019). With improvement in imaging and drug delivery technologies, theranostic nanoparticles are bound to be the main player in future cancer therapies, particularly in complex diseases like breast cancer.

V. Challenges and Limitations

Toxicity and Biocompatibility

Even as nanomedicine promises to make cancer treatments safer and more precise, concerns about the potential toxicity and long-term biocompatibility of nanoparticles remain among the most challenging hurdles to widespread clinical adoption. Although many

nanoparticles are engineered to be biocompatible—using materials like lipids, polymers, or metals such as gold—even these may exhibit unexpected behavior when subjected to biological environments over timescales that are long compared with chemical or physical scales. Their accumulation in critical organs such as the liver, spleen, or kidneys can lead to chronic toxicity or inhibit organ function. Additionally, nanoparticles could induce immune responses or activate the complement system, especially in hypersensitive individuals. Unforeseen side effects can occur due to their small size, which allows them to traverse biological barriers and penetrate spaces unavailable to larger molecules (Fadeel et al., 2013). These factors necessitate prolonged *in vivo* investigations, including pharmacokinetics, biodistribution, metabolism, and excretion profiles, to ensure the safe use of nanotherapeutics in cancer treatment.

Manufacturing and Scalability

One of the most significant technical hurdles in bringing nanoparticle treatments from the laboratory to the clinic is large-scale manufacturing. While many nanoparticle formulations are highly promising in preclinical models, large-scale production of these particles on a regular basis is full of many difficulties. Reproducibility of size, shape, surface charge, and functionalization is critical because even small differences can have a significant impact on the action of the particles in the body, e.g., biodistribution, drug release, and targeting efficiency. Moreover, synthesis of nanoparticles typically involves complex and time-consuming multi-step procedures, rigorous requirements of sterility, and high-tech equipment. Quality control and batch-to-batch reproducibility are a formidable challenge. These complexities raise the cost of production and slow the path to regulatory approval (Farokhzad & Langer, 2009). Developing scalable, stable, and cost-effective routes for fabricating nanoparticles is important to bridge the gap between research and therapeutic deployment.

Regulatory and Clinical Approval

The regulatory framework for nanomedicine is still in the process of being developed, and this uncertainty has emerged as a major bottleneck for the approval and availability of nanoparticle-based therapies. Unlike conventional drugs, nanoparticles exist in a regulatory gray area because they exhibit unique physicochemical properties and multivalency. Regulatory agencies like the U.S. Food and Drug Administration (FDA) require comprehensive safety and efficacy data, including detailed particle shape descriptions, release profiles, degradation products, and *in vivo* interactions. Because of their complex behavior in biological systems, typical toxicology models may not be adequate to examine their entire safety profile. Besides, regulatory systems across countries vary significantly, offering yet another complexity challenge for developers seeking to acquire global approvals. To date, only a handful of the nanomedicine products have obtained FDA approvals, emphasizing the gradual and laborious nature of the review process (Ventola, 2017). A more customized regulatory policy that considers the unique features of nanotherapeutics will be crucial in driving the industry forward.

Cost and Accessibility

The prohibitively expensive cost of producing and manufacturing nanoparticle-based drugs is another hurdle towards their widespread use, especially in low- and middle-income countries. The cost of raw material, sophisticated equipment, trained personnel, and compliance with regulations is extremely high. Furthermore, conducting prolonged preclinical and clinical trials to establish safety and efficacy adds to the cost. These economic burdens may make nano therapies significantly more expensive than conventional therapies, which could limit them to being sold to wealthier segments or nations that have advanced healthcare systems. Such disparity could contribute to expanding existing healthcare gaps, particularly in parts of the world where breast cancer rates are progressively rising but access to cutting-edge therapies remains limited. Until cost-reduction strategies and equitable distribution models are designed, the global implications of nanotechnology for breast cancer treatment can be maintained unrealized (Wagner et al., 2018).

VI. DISCUSSION

Passive Targeting

Passive targeting is perhaps the most rudimentary mechanism by which nanoparticles induce targeted drug delivery to tumor tissues. The method essentially leverages the EPR effect, a phenomenon well documented in solid tumors. Compared to normal tissue, tumors develop abnormally permeable vasculature due to their chaotic and rapid angiogenesis. These perivascular channels typically have wide fenestrations measuring between 100 nm and even several micrometers, through which nanoparticles extravasate passively into a tumor interstitium. At the same time, the inefficiency of lymphatic drainage within tumor regions results in poor clearance, resulting in prolonged residence of the particles within the tumor microenvironment (Maeda et al.).

The significance of the EPR effect lies in the fact that it is possible to transfer an element of selectivity to non-targeted forms of drugs. Traditional chemotherapeutic drugs diffused systematically throughout the organism, killing healthy and malignant cells alike, hence the disabling chemotherapy side effects. In contrast, nanoparticles designed to take advantage of the EPR effect can selectively accumulate in tumor tissue, essentially localizing the therapeutic payload where it is needed most without affecting normal cells.

After the nanoparticles have passively accumulated in the interstitium of the tumor, the second challenge is intracellular delivery. In order to be of therapeutic benefit, the drugs carried by these particles need not only to enter the tumor but to traverse cellular membranes efficiently. This is accomplished typically by endocytosis, whereby the cancer cell membrane engulfs the nanoparticle, bringing it inside to the cytoplasm. But internalization by endocytosis is only part of the process. Once within the cell, nanoparticles tend to be trafficked to endosomes and then lysosomes, where degradative enzymes and acidic environments can break down the therapeutic

agent. To maintain drug bioavailability and therapeutic efficacy, therefore, nanoparticles must be designed to escape the endosome prior to degradation.

Most nanoparticles today feature smart design elements that enable such endosomal escape. Others are built using pH-sensitive materials that destabilize at the acidic endosomal pH, releasing directly into the cytoplasm. Others employ the "proton sponge effect," wherein polymeric elements swell with proton accumulation, inducing disruption of the endosomal membrane and release of the payload (Sahay et al.). This type of engineering targets the release of the drug in exactly the proper location where it will be most effective: in the cytosol of the cancer cell.

Furthermore, receptor-mediated endocytosis has also been shown to be a valuable complement to passive targeting. Through the binding of ligands or antibodies on the nanoparticle surface that target receptors generally overexpressed on breast cancer cells—such as folate receptors, HER2, or transferrin receptors—nanoparticles can be internalized more specifically. By this two-stage mechanism, whereby EPR leads to localization and receptor-targeting leads to increased uptake, a highly specific mechanism for drug delivery into tumor cells can be realized.

The second notable benefit of endocytotic delivery by nanoparticle is avoidance of drug-resistance mechanisms. Multidrug resistance (MDR) of metastatic breast cancer and most cancers is normally due to efflux pump overexpression, including P-glycoprotein (P-gp), which functions actively to rid the cancer cell of chemotherapy. Nanoparticles, however, are delivered inside cells via vesicular transport and thereby do not get excreted out from the cell as a consequence of these pumps but are able to maintain inside the cell active concentrations of drug cargo.

In addition, nanoparticle-based delivery systems shield labile drugs from early degradation by protecting them during transit and releasing them only after reaching the cancerous cells. This not only enhances the stability of the drugs but also maximizes the therapeutic index. Such approved nanomedicine formulations as Doxil (liposomal doxorubicin) and Abraxane (albumin-bound paclitaxel) illustrate this principle. These agents exploit the EPR effect and passive targeting mechanisms to achieve improved tumor accumulation, prolonged circulation half-lives, and reduced off-target toxicity compared to their free-drug counterparts.

Together, passive targeting by the EPR effect and endocytosis-assisted internalization is an incredibly powerful collaborative effort in breast cancer drug delivery through nanoparticles. With scientists refining the physicochemical characteristics of these particles—tuning particle size, shape, surface chemistry, and charge—their ability to exploit tumor biology even more selectively will only grow, with a future that can potentially revolutionize the way we treat breast cancer chemotherapy and significantly enhance the chances of patients all over the world.

Active Targeting

Active targeting is a more specific and advanced method in nanoparticle-based drug delivery that is founded on passive targeting and further enhances the selectivity of therapeutic

delivery. In this case, nanoparticles are functionalized with biological entities such as ligands, monoclonal antibodies, or peptides that specifically target the overexpressed receptors or antigens on the surface of cancer cells. This selective binding facilitates receptor-mediated endocytosis, where the drug-loaded nanoparticle is engulfed by the target cell. The aim is to achieve maximum drug concentration in cancerous tissues and minimum exposure to normal cells, thereby circumventing systemic toxicity and increasing therapeutic efficacy.

Breast cancer is a very heterogeneous cancer, classified into various subtypes based on the presence of certain molecular markers. The most common subtypes are hormone receptor-positive tumors, with ER and/or PR expression, HER2-overexpressing tumors with elevated human epidermal growth factor receptor 2 (HER2) expression, and triple-negative breast cancer (TNBC) with no ER, PR, or HER2 expression. Of all these, HER2-overexpressing breast cancer is most virulent and accounts for about 15–20% of total breast cancer diagnoses. Historically associated with a poor prognosis, HER2-overexpressing tumors today are now an initial focus of antibody-based therapies due to their distinct molecular profile and sensitivity to HER2 inhibition.

An excellent example of active targeting is trastuzumab (Herceptin)-conjugated nanoparticle conjugates, which is a monoclonal antibody that selectively targets the HER2 receptor. These HER2-targeting nanoparticles have a dual function: they enable direct delivery of chemotherapeutic agents—such as paclitaxel or doxorubicin—into HER2-overexpressing tumor cells, and also inhibit downstream HER2-mediated tumor-promoting survival signals. The dual mechanism enhances drug delivery to the tumor and oncogenic signal suppression, ultimately offering increased clinical benefits with reduced systemic toxicity (Park et al.).

Along with antibody-targeting, peptide-mediated delivery systems have also shown promise. A case in point is the delivery by means of RGD peptides, which are fibronectin-derived short peptides of high affinity for binding to integrins, particularly overexpressed $\alpha v \beta 3$ integrins on the tumor angiogenic endothelial cell surface. These RGD-functionalized nanoparticles increase intracellular uptake and tissue penetration by binding to the tumor vasculature and inducing transcytosis across endothelial barriers. For breast cancer, especially in highly aggressive tumors with extensive angiogenesis, this strategy can enhance the therapeutic index of anticancer drugs and improve access to hypoxic tumor cores that are otherwise inaccessible by passive diffusion alone.

Another effective approach is ligand-based targeting, whereby nanoparticles are functionalized with folic acid or transferrin as ligands to target receptor-mediated endocytosis in cancer cells. The folate receptor (FR) is overexpressed in many aggressive breast cancer subtypes and is an important target for folic acid-conjugated nanoparticles. Also, the transferrin receptors (TfRs) on proliferating cancer cells are overexpressed due to these cells' increased demands for iron, and transferrin-conjugated nanoparticles have been engineered to exploit this overexpression. The nanoparticles bind to such receptors and can be delivered into the cancer cells by an efficient process of internalization, bypassing drug resistance phenomena and non-specific uptake by normal cells (Kobayashi et al., 2014). This approach has proven

particularly useful in avoiding multidrug resistance and increasing drug accumulation within resistant tumor phenotypes.

Also, active targeting enables the integration of diagnostic imaging and therapeutic delivery in a single step, which has been termed as theranostics. Through the labeling of target nanoparticles with contrast agents or fluorescent probes, scientists can monitor biodistribution and deposition of the nanoparticles in real time, so clinicians can monitor treatment response and adjust therapy protocols. This is revolutionary for precision medicine, which enables dynamic alteration of treatment based on the patient's own tumor nature and response.

Stimuli-Responsive Delivery

Stimuli-responsive delivery systems are a new advancement in the field of nanoparticle engineering, where drug release is induced depending on certain conditions of the tumor microenvironment. The systems are designed to react to physiological cues such as acidic pH, elevated temperature, enzymatic action, redox gradients, or hypoxia—conditions that are typically present in the microenvironment of cancerous tissues but not under normal tissues.

Among the most widely explored stimuli-responsive approaches is pH-sensitive delivery. Tumor tissues generally possess a weakly acidic extracellular pH (approximately 6.5–6.8), compared with the almost neutral pH of normal tissues (approximately 7.4). Additionally, endosomal and lysosomal cellular compartments possess even lower pH levels. Nanoparticles that swell or degrade under acidic conditions can thus deliver their therapeutic payloads selectively at the tumor or intracellularly inside the cell's intracellular compartments. The pH-responsive release mechanism minimizes leakage of the drug upon circulation and provides delivery of maximum drug concentration to the site of need (Zhang et al., 2019).

Temperature-sensitive nanoparticles are another lovely strategy. They remain stable at physiological temperatures (37°C) but are triggered to alter their structure at slightly elevated temperatures (around 42°C), which can be initiated by external heat application techniques such as concentrated ultrasound or laser irradiation. This heat-sensitive activation could potentially impact the rapid and focal release of drugs within enclosed spaces in tumor tissues, especially when coupled with hyperthermia-treated treatments.

Enzyme-responsive nanoparticles exploit the overexpression of specific enzymes in the tumor microenvironment, for example, matrix metalloproteinases (MMPs), cathepsins, or phospholipases. The enzymes cleave specific substrates or bonds within the nanoparticle matrix, triggering the drug release in enzyme-rich environments, thereby enhancing spatial precision.

Stimuli-responsive nanoparticles, particularly with the addition of active targeting moieties, offer a very sophisticated level of control over drug release and delivery. This multi-modal approach reduces the likelihood of premature drug activation and limits systemic toxicity. Moreover, by responding to internal or externally administered triggers, these systems allow flexibility for localized and systemic treatment approaches.

Generally speaking, the synergy of active targeting and stimulus-sensitive delivery into nanomedicine formulations is an important breakthrough in the treatment of breast cancer. Taken

together, all these technologies heighten the specificity, efficacy, and safety of therapeutic protocols, promising new hope to patients with difficult-to-treat subtypes and aligning with the grand goals of precision oncology.

VII. CONCLUSIONS

Nanotechnology has emerged as a groundbreaking advancement in the field of breast cancer therapy, offering the potential for more targeted, efficient, and personalized treatments. The use of nanoparticles—such as liposomes, polymeric nanoparticles, gold nanoparticles, and carbon nanotubes—has revolutionized the delivery of chemotherapeutic agents, reducing systemic toxicity while improving drug bioavailability and therapeutic efficacy.

The two primary approaches to nanoparticle drug delivery—passive targeting through the EPR effect and active targeting using specific ligands, antibodies, or peptides—have significantly enhanced the precision of breast cancer treatment, ensuring that drugs are delivered directly to cancer cells and minimizing damage to healthy tissues. Additionally, stimuli-responsive delivery systems offer further refinement, enabling the release of drugs in response to specific tumor microenvironmental conditions, further increasing the specificity and safety of the treatment.

Despite these promising advancements, several challenges remain. Concerns regarding the toxicity and biocompatibility of nanoparticles, as well as the high cost of their production, continue to pose barriers to widespread clinical adoption. Additionally, the slow pace of regulatory approval for nanoparticle-based therapies highlights the need for continued research to address safety and efficacy concerns.

Nonetheless, the potential of nanotechnology to improve breast cancer treatment is vast. As research continues to progress, it is expected that nanoparticle-based therapies will become an integral part of breast cancer care, offering patients more effective treatments with fewer side effects. Continued focus on optimizing nanoparticle formulations, ensuring their biocompatibility, and expediting clinical translation will pave the way for more accessible and personalized cancer therapies, ultimately improving the outcomes for breast cancer patients worldwide.

WORKS CITED

- ((Barenholz, Y. (2012). Liposome applications: Problems and prospects. *Cancer Research*, 72(12), 2010–2017.)) Barenholz, Y. (2012). Liposome applications: Problems and prospects. *Cancer Research*, 72(12), 2010–2017.
- ((Chen, Q., Feng, L., & Li, C. (2016). Nanoparticle-based approaches to combat multidrug resistance in cancer therapy. *Nature Nanotechnology*, 11(6), 542–553.)) Chen, Q., Feng, L., & Li, C. (2016). Nanoparticle-based approaches to combat multidrug resistance in cancer therapy. *Nature Nanotechnology*, 11(6), 542–553.
- ((Danhier, F., Feron, O., & Preat, V. (2012). To exploit the tumor microenvironment: Passive and active tumor targeting of nanocarriers for anticancer drug delivery. *Journal of Controlled Release*, 148(2), 135–146.)) Danhier, F., Feron, O., & Preat, V. (2012). To exploit the tumor microenvironment: Passive and active tumor targeting of nanocarriers for anticancer drug delivery. *Journal of Controlled Release*, 148(2), 135–146.
- ((Fadeel, B., & Garcia-Bennett, A. (2013). Better safe than sorry: Understanding the toxicological properties of nanomaterials. *Advanced Drug Delivery Reviews*, 65(1), 101–113.)) Fadeel, B., & Garcia-Bennett, A. (2013). Better safe than sorry: Understanding the toxicological properties of nanomaterials. *Advanced Drug Delivery Reviews*, 65(1), 101–113.
- ((Farokhzad, O. C., & Langer, R. (2009). Impact of nanotechnology on drug delivery. *ACS Nano*, 3(1), 16–20.)) Farokhzad, O. C., & Langer, R. (2009). Impact of nanotechnology on drug delivery. *ACS Nano*, 3(1), 16–20.
- ((Gao, X., Yang, L., & Sun, W. (2005). Quantum dots for in vivo molecular imaging. *Advanced Drug Delivery Reviews*, 57(1), 35–46.)) Gao, X., Yang, L., & Sun, W. (2005). Quantum dots for in vivo molecular imaging. *Advanced Drug Delivery Reviews*, 57(1), 35–46.
- ((Huang, X., Zhang, F., & Li, L. (2011). Gold nanoparticle-based photothermal therapy for breast cancer treatment. *Nanotechnology*, 22(21), 215104.)) Huang, X., Zhang, F., & Li, L. (2011). Gold nanoparticle-based photothermal therapy for breast cancer treatment. *Nanotechnology*, 22(21), 215104.
- ((Jiang, Y., Liu, J., & Wu, Y. (2019). Theranostics: Combining diagnostic and therapeutic functions with nanotechnology. *Advanced Drug Delivery Reviews*, 142, 121–131.)) Jiang, Y., Liu, J., & Wu, Y. (2019). Theranostics: Combining diagnostic and therapeutic functions with nanotechnology. *Advanced Drug Delivery Reviews*, 142, 121–131.
- ((Kobayashi, Hiroto, et al. "Targeted delivery of anticancer drugs with folate and transferrin receptor-mediated endocytosis." *Advanced Drug Delivery Reviews*, vol. 66, 2014, pp. 26–39.)) Kobayashi, Hiroto, et al. "Targeted delivery of anticancer drugs with folate and transferrin receptor-mediated endocytosis." *Advanced Drug Delivery Reviews*, vol. 66, 2014, pp. 26–39.
- ((Liu, Z., & Robinson, J. T. (2008). Carbon nanotubes for drug delivery: A novel class of drug carriers. *Advanced Drug Delivery Reviews*, 60(11), 1054–1063.)) Liu, Z., & Robinson, J. T. (2008). Carbon nanotubes for drug delivery: A novel class of drug carriers. *Advanced Drug Delivery Reviews*, 60(11), 1054–1063.

- ((Maeda, Hiroshi, et al. "Tumor vascular permeability and the EPR effect in macromolecular therapeutics: a review." *Journal of Controlled Release*, vol. 65, no. 1–2, 2000, pp. 271–284.)) Maeda, Hiroshi, et al. "Tumor vascular permeability and the EPR effect in macromolecular therapeutics: a review." *Journal of Controlled Release*, vol. 65, no. 1–2, 2000, pp. 271–284.
- ((Park, Jong W., et al. "Anti-HER2 Immunoliposomes: Enhanced Efficacy Attributable to Targeted Delivery." *Clinical Cancer Research*, vol. 8, no. 4, 2002, pp. 1172–1181.)) Park, Jong W., et al. "Anti-HER2 Immunoliposomes: Enhanced Efficacy Attributable to Targeted Delivery." *Clinical Cancer Research*, vol. 8, no. 4, 2002, pp. 1172–1181.
- ((Peer, D., Karp, J. M., & Hong, S. (2007). Nanocarriers as an emerging platform for cancer therapy. *Nature Nanotechnology*, 2(12), 751–760.)) Peer, D., Karp, J. M., & Hong, S. (2007). Nanocarriers as an emerging platform for cancer therapy. *Nature Nanotechnology*, 2(12), 751–760.
- ((Rafiq, Muhammad, et al. "Recent Advancements in Nanotechnology-Based Breast Cancer Therapy." *NanoImpact*, vol. 30, 2024, article no. 100524,.)) Rafiq, Muhammad, et al. "Recent Advancements in Nanotechnology-Based Breast Cancer Therapy." *NanoImpact*, vol. 30, 2024, article no. 100524,.
- ((Sahay, Gaurav, et al. "Endocytosis of nanomedicines." *Journal of Controlled Release*, vol. 145, no. 3, 2010, pp. 182–195..)) Sahay, Gaurav, et al. "Endocytosis of nanomedicines." *Journal of Controlled Release*, vol. 145, no. 3, 2010, pp. 182–195..
- ((Tiwari, G., Tiwari, R., & Ghosh, V. (2012). Nanoparticles as drug delivery systems: Potential and challenges. *Drug Development and Industrial Pharmacy*, 38(10), 1052–1064.)) Tiwari, G., Tiwari, R., & Ghosh, V. (2012). Nanoparticles as drug delivery systems: Potential and challenges. *Drug Development and Industrial Pharmacy*, 38(10), 1052–1064.
- ((Ventola, C. L. (2017). Progress in nanomedicine: Approved and investigational nanodrugs. *P&T*, 42(12), 798–809.)) Ventola, C. L. (2017). Progress in nanomedicine: Approved and investigational nanodrugs. *P&T*, 42(12), 798–809.
- ((Wagner, V., Dullaart, A., & Bock, A. (2018). The emerging nanomedicine landscape. *Nature Nanotechnology*, 13(6), 453–465.)) Wagner, V., Dullaart, A., & Bock, A. (2018). The emerging nanomedicine landscape. *Nature Nanotechnology*, 13(6), 453–465.
- ((Zhang, Rui, et al. "Stimuli-Responsive Nanocarriers for Therapeutic Applications in Cancer." *Theranostics*, vol. 9, no. 12, 2019, pp. 3249–3271. doi:10.7150/thno.33784.)) Zhang, Rui, et al. "Stimuli-Responsive Nanocarriers for Therapeutic Applications in Cancer." *Theranostics*, vol. 9, no. 12, 2019, pp. 3249–3271. doi:10.7150/thno.33784.

Results

Methods

This paper was developed using a literature review methodology. A range of peer-reviewed scientific papers, clinical reports, and journal articles on nanotechnology and breast cancer treatment were analyzed. Sources were selected based on their relevance, credibility, and recency, with an emphasis on studies involving nanoparticle drug delivery systems. No new experiments were conducted; the synthesis presented reflects current knowledge and ongoing clinical and preclinical research.

Can GPT Generated Poetry, Infused with Extreme Metaphors Be Perceived as More Profound than Original Human Written Poetry? By Devansh Gupta

Abstract

Recent technological progress in the field of Natural Language Processing has equipped Large Language Models such as GPT with the ability to generate poetry that closely resembles human authored poetry in terms of literary style, imagery, and emotional tone. This study examines whether such GPT generated poetry that are infused with such extreme metaphors are perceived more profoundly than original human written poetry. The study uses a blind comparative reader response methodology to evaluate four poems – a purely human authored poem, a purely AI authored poem infused with extreme metaphors, and two hybrid poems created through human-AI and AI-human refinement. The study reveals that though human generated poetry elicits stronger emotional impact, GPT generated poetry is admired for its bold imagery and conceptual originality. These findings have a profound implication on future literary evaluation, and for understanding how AI generated literature may reshape aesthetic standards and reader expectations.

Keywords

AI generated poetry, reader response theory, human AI hybridity, literary aesthetics

1. Introduction

In recent years, the rise of artificial intelligence in the creative writing industry has immensely shaped how we respond to literature and poetry. Advances in natural language processing, notably through models such as OpenAI's Generative Pre-trained Transformer (GPT), have enabled machines to produce texts that closely mimic human writing styles and emotional connect. Research by Köbis and Mossink (106553) shows that readers frequently fail to differentiate AI-authored poems from those composed by humans. This challenges long standing assumptions regarding creativity, originality, and authenticity in literary production, thereby forcing us to reconsider the boundaries between what is considered "human" and "AI".

Metaphors have held an essential part in poetry, enabling readers to grasp abstract ideas by connecting them with more familiar imagery. However, not all metaphors function on the same cognitive or emotional level. There exists a spectrum, and within this spectrum, *extreme metaphors* stand out to be very potential. Unlike ordinary metaphors, such as "time is a thief" or "her voice was music", which rely on familiar symbolic associations, extreme metaphors stretch boundaries of logic and association. They often link two seemingly unrelated things, requiring readers to engage and interpret with the work. .

For example, an extreme metaphor like "*grief is a cathedral built in reverse, with stained glass shattering inward*" makes very deep logical and emotional associations. It contains many symbolic layers - construction, fragility, religion, to evoke complex emotional states. Such metaphors call for higher-order semantic integration. Research by Stamenković, Ichien, and

Holyoak (285–301) highlights how the comprehension and appreciation of such metaphors depend heavily on context, intellect, and language. Extreme metaphors elevate the impact of a piece of poem by forcing more engagement from the reader.

This study addresses the research question of how extreme metaphors in AI-generated poetry influence reader perception in comparison to those in human-generated works. Specifically, it seeks to explore whether the complexity of metaphors vary and whether metaphors inherent in AI-generated poems can contain depth, meaning, and impact compared to human written poetry. The objectives include a comparative analysis of reader responses to selected poems, assessment of engagement with content, and evaluation of such implications.

The novelty of this research cannot be understated. This study provides a foundation for understanding how AI-generated literature reshapes the criteria of how poetry is evaluated.

2. Literature Survey

With the sophistication of AI in poetry, a lot of research has been conducted into how machine-generated texts are perceived by human audiences. Central to this discourse is the question of whether AI, particularly models like ChatGPT, can evoke the same sense of emotion traditionally associated with human-written poetry.

Köbis and Mossink (2020) investigated whether readers could distinguish between human and GPT-2-generated poetry. Participants often failed to identify the AI origin and sometimes even preferred the algorithmic poems. Their study suggests that AI can convincingly mimic human poetic styles, especially when guided by human selection.

Porter and Machery (2024) extended this inquiry by comparing GPT-3 and GPT-4 poems with canonical poetry. Participants rated AI poems higher across various emotional and aesthetic dimensions, including beauty, rhythm, and meaning. Interestingly, many attributed AI poems to human authors, revealing a “more human than human” bias.

Stamenković, Ichien, and Holyoak (2020) approached metaphor from a logical standpoint, analyzing how individuals comprehend context based metaphors. Their findings show that complex metaphors - those requiring higher-level understanding- were more likely to be perceived as meaningful when adequately framed.

Finally, **Ichien, Stamenković, and Holyoak (2023)** demonstrated that GPT-4 can interpret novel literary metaphors with striking insight, sometimes exceeding the interpretive depth of human participants.

3. Methodology and Analysis

3.1 Methodology

This study uses a comparative reader-response methodology to investigate whether GPT-generated poetry infused with extreme metaphors can be perceived as more profound than human-authored poetry. The design uses blind evaluation, ensuring that participants’ judgments

are based purely on intent and emotional resonance rather than prior knowledge of AI or human authorship. Such an approach helps minimise the bias present while choosing a piece of work.

3.2 Poem Selection

Four poems were selected:

1. **Poem A (Human)** – Purely human-authored poem with conventional lyrical structure.
2. **Poem B (AI)** – Fully GPT-generated poem employing extreme metaphors
3. **Poem C (Human written and AI refined)** – Human-authored poem refined by AI
4. **Poem D (AI written and Human refined)** – AI-generated poem refined by a human

This helped analyse the direct comparison of pure human, pure AI, and hybrid forms.

3.3 Participants

The survey drew responses from 145 participants spanning diverse ages and qualifications.. Ages broke down as follows: 13.9% (n=20) (in the paper, n refers to number of participants) below 15, 24.3% (n=35) aged 15–30, 58.3% (n=84) aged 30–50, and 3.5% (n=5) above 50.

Education levels included 62.5% (n=90) postgraduates, 19.4% (n=28) undergraduates, 10.4% (n=15) with high school or below, and 7.7% (n=11) holding other professional degrees such as MBA or B.Ed. Poetry familiarity varied: 11.1% (n=16) literati enthusiasts, 33.3% (n=48) regular readers, 43.1% (n=62) occasional readers, and 12.5% (n=18) rare or non-readers.

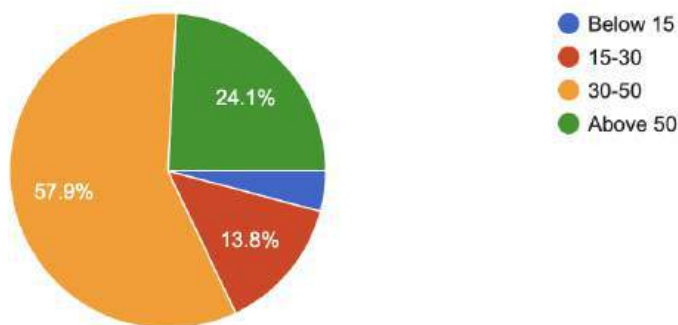


Figure 1: Age demography of responses collected on survey

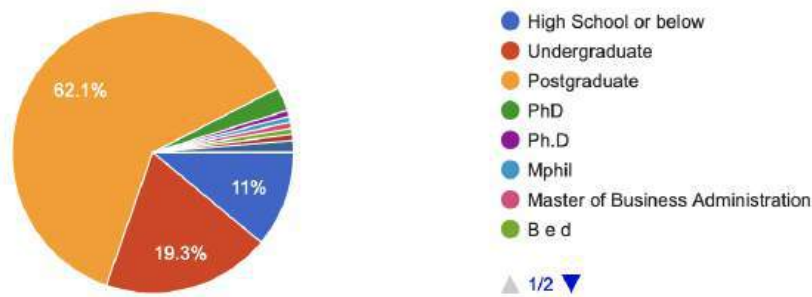


Figure 2: Qualification demography of responses collected on survey

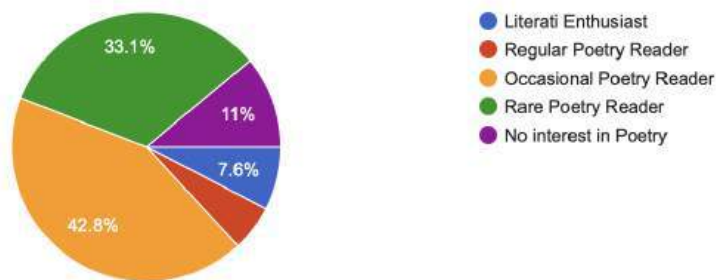


Figure 3: Poetic interest demography of responses collected on survey

3.4 Survey Procedure

Participants viewed the four poems in random order via a common Google Form and rated them across four categories: the most beautiful metaphor usage, emotional resonance, perceived human authorship, and perceived AI authorship. No hints about authorship were given.

3.5 Questionnaire

The poems selected were as follows:

Poem A

Clouds come floating into my life,
no longer to carry rain or usher storm,
but to add color to my sunset sky

Source - Rabindranath Tagore, Stray Birds

Poem B

Grief is a staircase of mirrors I climb,

each step breaking until my reflection
shatters into rain across the sky
Source - ChatGPT

Poem C

The park is quiet after rain,
puddles mirror passing feet;
a sparrow hops along the path,
and the morning feels complete
Source - Devansh Gupta + ChatGPT

Poem D

The sky is dusty gray as
rivers drag their plastic down;
the air grows heavier each day,
and flowers fade within the town

Source - ChatGPT + Devansh Gupta

4. Analysis and Implications

4.1 Emotional Impact

Poem A (human) got the highest emotional response at 40 % (n=58), ahead of pure AI poems. Hybrid Poem C held strong at 22.1% (n=32), showing AI edits on human work don't kill the emotion. Poem B (AI with extreme metaphors) had a moderate score at 18.6% (n=27). Poem D (AI edited by human) scored lowest at 12.4% (n=18), meaning human fixes keep it readable but don't add heart. People aged 30-50 and postgrads rated human poems as more moving. Younger (15-30) and casual readers split votes between AI and hybrids, relying on fresh metaphors over subtle emotions.

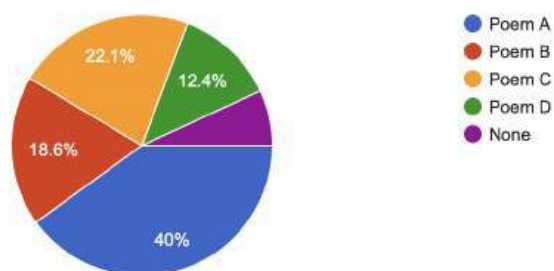


Figure 4: Emotional resonance demography of responses collected on survey

4.2 Metaphorical Appreciation

For "most beautiful metaphor": Poem A (human written) led at 35.9% (n=52), topping Poem B (AI) at 26.9% (n=39). Hybrids trailed behind: Poem C at 18.6% (n=27) and Poem D at 15.2% (n=22). This shows extreme metaphors do not add extra boosts to AI's sense of depth. Refining them (by AI or human) makes them feel less striking. Casual readers and ages 15-30 loved Poem B's boldness. Experts and postgrads picked Poem A, preferring classic human style.

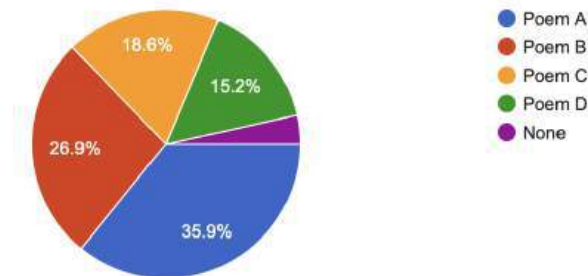


Figure 5: Metaphor appreciation demography of responses collected on survey

4.3 Perceived Human Authorship:

Votes for "most human-written": Poem A got 40% (n=58), Poem B 13.1% (n=19), Poem C 27.6% (n=40), Poem D 16.6% (n=24). Poem C (human + AI) felt reliably human. Poem D (AI + human) often seemed AI. Younger and casual readers misjudged Poem B most. Postgrads spotted AI better.

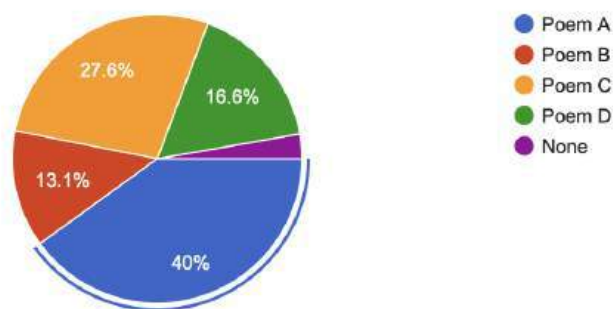


Figure 6: Perceived human authorship demography of responses collected on survey

5. Perceived AI Authorship

For "most AI-generated": Poem D topped at 42.4%, Poem B at 25.7%, Poem C at 16.7%, Poem A at 15.3%. Poem D's machine-like writing was perceived the same even with human edits

. Poem B hid its AI side moderately well with metaphors. Experts and regular readers caught AI easiest. Younger casual readers bought Poem B as human, distracted by its flashy ideas.

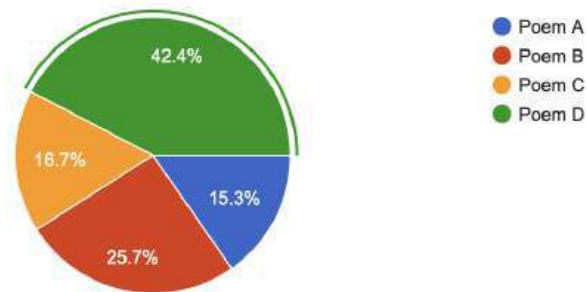


Figure 7: Perceived AI authorship demography of responses collected on survey

6. Conclusion

This study suggests GPT-written poems can *seem* as deep as human poems when they use bold, extreme metaphors. Readers can recognise clever imagery and strong ideas in AI poetry almost as easily as in human writing. But emotional impact is where humans still win—people feel more personally moved by human poems, even if they admire AI’s technique.

The ability to spot AI also depends on the reader. Younger readers with little poetry experience often assume AI poems are human-made, especially when the metaphors are intense. Older, well-read readers identify AI more accurately, likely because they notice patterns that feel less organic.

Overall, the findings show that how we judge poetry may be shifting: instead of linking “depth” only to human creativity, readers may increasingly focus on originality, complexity, and metaphor—whether the poem is written by a person, an AI, or both.

Works Cited

- Ichien, Nicholas, Dušan Stamenković, and Keith J. Holyoak. "GPT-4's Comprehension and Interpretation of Novel Literary Metaphors." *Proceedings of the Annual Meeting of the Cognitive Science Society*, 2023.
- Köbis, Nils, and Lars D. Mossink. "Artificial Intelligence versus Maya Angelou: Experimental Evidence That People Cannot Differentiate AI-Generated from Human-Written Poetry." *Computers in Human Behavior Reports*, vol. 1, 2020, p. 100005. ScienceDirect, doi:10.1016/j.chbr.2020.100005.
- Porter, Ethan, and Edouard Machery. "ChatGPT Can Write Better Poetry Than Humans." *Nature Human Behaviour*, 2024.
- Stamenković, Dušan, Nicholas Ichien, and Keith J. Holyoak. "Metaphor Comprehension: An Individual-Differences Approach." *Journal of Memory and Language*, vol. 112, 2020, pp. 104087. Elsevier, doi:10.1016/j.jml.2020.104087.
- Tagore, Rabindranath. *Stray Birds*. Macmillan, 1916.

Leveling the Playing Field: A Comparison of Economic Structures and Systems Across Different Sports Leagues By August Berglind

Abstract

The purpose of this paper is to make the reader and sports fans aware of the economics behind different sports and leagues. This includes the history, system, and benefits and detriments to the system of four different leagues. There is also a comparison section, which compares the differences between the four different leagues. This comparison section also compares leagues and systems to wider economic ideas and beliefs that exist outside of sports. The research done has shown that each system has its own good and bad ideas, and that there is no perfect way to do this. The paper also shows what if more time and the ability to collect data that is not already out on the interwebs could help answer some unanswered ideas. The paper has concluded that sports leagues are very economically vast and different, and that having an understanding of how they work is essential to the viewer.

Keywords Sports, Economics, Leagues

Introduction

Many people across the world watch sports, with over 3.5 billion watching sports in some form. There are also plenty of different kinds of sports, and those sports run on different systems across the world. For the purpose of this paper, there are two main leagues- the National Hockey League which represents North American sports as it is the most “average” system of the big four sports in North America, as it doesn’t have any weird quirks like the other three. The other main league is the English Premier League, which represents European football. The ROSHN Saudi League and NCAA College Football are also included but less important due to how unique they are compared to other leagues across the world. This paper sets out to answer questions about economic differences between leagues. Plenty watch these leagues without a true understanding of what goes on beyond the TV. This paper will cover key components and vocabulary that is used to separate the leagues. It will then cover a basic history of each league, and then will talk about its economic system and the pros and cons of their respective system. While the four leagues chosen are evenly spread out to give a variety of different types of sporting systems, it is important to note that no two leagues are the same. For example, the NHL represents North America as I consider it to be the most “basic” system of the big four sports leagues in North America. The MLB has no salary cap at all, which shows that despite the NHL being the chosen league as the middle ground between all four major North American Leagues, the NHL still has significant variation with other North American sports. These 4 leagues chosen do not represent everything, they are just there as a baseline.

League Components: Important Considerations for Leagues

Open League

One of the most important concepts regarding league structures is open vs closed league formats. In open structures, leagues, anyone is allowed to start a team, who through promotion can make the top tier. These are called open leagues. Promotion and relegation is a concept popular in Europe that rewards teams who do well with promotion to a higher league, whereas teams who perform poorly get relegated to a lower league. This is a much more capitalistic idea, as you are rewarded on merit, as opposed to the North American league system where you are rewarded for doing poorly to try and keep everyone equal.

Closed League

A closed league means that every team has to start at the top level, with no promotion or relegation. An expansion fee is almost always necessary, along with all other owners of franchises voting to allow these new teams. These closed leagues have drafts, where young players who are getting ready to play in the league register, get chosen by a team, and that team has their contract rights. Closed leagues primarily exist in North America.

Collective Bargaining Agreements (CBAs)

CBAs (collective bargaining agreements) is a term, primarily in North American sports, used to set outlines for things such as the salary cap, player pay, and other various things like playoff structure, which is agreed on by a mix of players and the league.

Player Movement

Free agency is a term in closed leagues for when a player's contract runs out of term, they are able to sign with any team they want, but with some restrictions and players are eligible to sign extensions before their contract runs out. Transfers are a similar idea in open leagues, except players can transfer while on contract if both teams approve it. Transfers are also usually paid, so teams sell players for money if there is a mutual want to sell and buy them. Players can also be swapped for each other in both closed and open leagues, considered a swap in open leagues and a trade in closed leagues. There is much more power to the player in open leagues.

Salary Caps & Finances

Closed leagues have CBAs and the ability to have a salary cap, something not present in open leagues. A salary cap determines how much money a team can spend on players. There are hard caps, which provide zero ways to get around the cap, and soft caps, which allow there to be ways to get around the cap, but usually with penalties which increase with how much over the cap you are. Most open leagues do have monetary restrictions, which in England is called FFP (Financial Fair Play), which is a much softer version of a salary cap intended to prevent teams from just spending as much money as possible to create the best teams which would ruin competitiveness. The major sources of money for the sports leagues compared are ticket sales; which is how many people buy tickets to attend the game, sponsorships; which is companies paying the sporting clubs and leagues to be able to have their logo in things like billboards,

events or jerseys, merchandising; which involves the sale of things such as scarves, jerseys and hats, and TV contracts, which involves the leagues giving an outside organization the right to broadcast games for a sum of money.

The National Hockey League

History of the National Hockey League

The NHL is North America's premier ice hockey league, founded in 1917, across the Northeastern USA and Southeastern Canada. It started with five teams across Ontario and Quebec, being the Montreal Canadiens, Montreal Wanderers, Ottawa Senators, Quebec Bulldogs, and Toronto Arenas, with the Bulldogs opting out of the first season. The Stanley Cup is the most significant trophy in both the league and the sport. In the 1925-26 season, the Boston Bruins joined the league, marking the first American team. Chicago, Detroit and New York got teams in the 1926-27 season. By 1942-43, there were only six teams in the league. The Boston Bruins, the Chicago Blackhawks, the Detroit Red Wings, the Montreal Canadiens, the New York Rangers and the Toronto Maple Leafs. These teams existed for 24 years as the only six teams in the league, and are referred to as the "Original Six". In 1967-68, six new teams were added, the California Seals, Los Angeles Kings, Minnesota North Stars, Philadelphia Flyers, Pittsburgh Penguins and St. Louis Blues. This marked an expansion into Western North America that did not previously exist and brought the league-wide total to twelve teams. By the end of the 1970's, the NHL had 21 total teams. In 1979, the NHL added 4 teams from its merger with the WHA, and it was this merger which brought Wayne Gretzky and the Edmonton Oilers to the NHL, and Gretzky is widely considered the best hockey player and one of the greatest athletes of all time. The 3 other teams added were the Quebec Nordiques, Winnipeg Jets, and Hartford Whalers. None of these teams other than the Oilers are in their original location, with relocations to Colorado, Utah and North Carolina respectively. There is however a current team in Winnipeg named the Jets, but is an entirely different franchise. By the 2000-01 season, there were 30 teams. There are currently 32 teams, with the Vegas Golden Knights being added in the 2017-18 season, and the Seattle Kraken in the 2021-22 season. The Montreal Canadiens have the most Stanley Cups with 24, followed by the Toronto Maple Leafs, Detroit Red Wings, Boston Bruins, and Chicago Blackhawks, showing the dominance of the original six. The Maple Leafs have notably not won since 1967, and a Canadian team has not won since Montreal in 1993.

Organization & Structure

The NHL is also the largest and most successful hockey league across the world. It is a closed league which has a CBA, and follows most traditional North American sports setup. It is the 4th biggest league in North America by viewership. It has a hard salary cap, meaning that the salary limit is strict and there are no things such as the luxury tax in the NBA which allows teams to go over that limit. It is a feeder system, as the AHL, the second tier league in North America feeds into NHL, ECHL feeds into AHL, and so on. The only significant trophy is the

Stanley Cup, which is given to the winner of the Stanley Cup Playoffs. Sixteen teams, 8 from each conference, make the playoffs, with 7 games series determining who advances to the next round. 44% of the league's revenue comes from seating sales, 19% comes from national media, 14% comes from sponsorships, 12% from local media and 12% from concessions, parking and other sources. Significantly more reliant on ticket sales than any other big 4 North American sports league.

Benefits & Detriments of League Structures

The NHL runs on the traditional North American League setup. While there are differences between the NHL and the NFL, NBA, and MLB, they all follow the same basic principal template. The advantages of the NHL are as following; parity, salary cap, lack of foreign competition, and lower feeder leagues for players who are not ready yet. The NHL has parity due to its salary cap and draft process, which ensures that one team will never be good for an extreme amount of time, and no team will forever be bad. The salary cap helps prevent rich owners from just buying any player they want and also keeps the balance of the league more even. The lack of foreign competition means that all of the best hockey players in the world play in the NHL, making the overall quality of the league extremely high. Lower feeder leagues like the AHL and ECHL within organizations allows teams to develop players who are not ready for the league. There are however some cons within this. One of these is that it is very hard to start a team. Due to expansion fees and having to be approved by all 32 owners, it makes starting a club from scratch impossible without a ton of funding. Secondly, all teams have to operate on the same success philosophy, which is either being bad and then using draft picks to either draft a team or trade for players. In other sports, clubs make their money by developing and selling players to another team. Lastly, the fact that all teams are owned by rich groups and located in big cities makes it so the fanbase and atmosphere is not like what is seen in leagues that are closed.

English Premier League (EPL)

History of the English Premier League and of the FA (Football Association)

The English Premier League is the top tier of English football, and generally regarded as the best league in the world. The EPL was founded in 1992, however the FA, the group that controls football in England, was founded in 1863. The FA was formed at the Freemasons' Tavern in Great Queen Street in London. The premise of the formation was for football to have a standard set of rules, much like cricket did. The rules were significantly different from how they are today. On July 20th, 1871 the FA cup was founded. This is considered one of the most historic tournaments in sports, and any club in England can win it, regardless of division or league. The FA in 1885 made it so that players could be paid professionally, as opposed to being amateurs before. The FA cup is a bracket style tournament, however the other significant trophy,

the trophy given to the winners of the top level, does not have a playoff and the team that ends with the most points wins the league.

Currently, each team plays 38 games, with a win being 3 points, a draw (tie) being one point, and a loss being worth 0. Fast forward to 1992, The Premier League was founded. This replaced the English Football League and the EPL became the top footballing division in England, however the trophies for the two are usually counted together. The teams who currently place in the top four qualify for the UEFA Champions League, where they play a tournament against other European teams, and this trophy is considered the holy grail for club football. The Champions league is a mix of group stage and knockout games, culminating in a one game final at a neutral site to decide the trophy, with other knockout ties being both a home and away game. The UCL (UEFA Champions League) also started in 1992, but the European cup started in 1955. Just like the EPL and EFL, the UCL and European cup are considered the same trophy. The FA Cup, Premier League and UCL is often called the treble, and the only two teams that have won the treble are 1998-99 Manchester United and 2022-2023 Manchester City.

Famous clubs include the two already mentioned, Liverpool, Arsenal, Chelsea, Tottenham, among many others. Manchester United and Liverpool have the most English titles with 20 each, followed by Arsenal at 13 and then Manchester City with 10. Arsenal have the most FA cups with 14, followed by Manchester United with 13, and then Liverpool, Chelsea and Tottenham Hotspur all tied with 8. Among English teams, Liverpool has the most UCL's with 6, followed by Manchester United with 3, and then Chelsea and Nottingham Forest with 2. Manchester City and Aston Villa have also won one singular UCL. Famous players in both the EPL and EFL all time include Thierry Henry, Cristiano Ronaldo, Harry Kane, Alan Shearer, Kevin De Bruyne, Frank Lampard along with many others.

Organization and Structure

The FA and EPL operate on an open system, meaning anyone can start or own a team and can progress its way up through promotion and relegation. There is neither a CBA or salary cap in the league. Instead of free agency, they have transfers. There are FFP (Financial Fairplay) rules which help enforce that no team just spends as much money as they would like on players. The EPL, unlike the NHL, does not have a monopoly on the sport. The most significant sources of revenue are broadcasting rights and commercial revenue. There are also transfer windows which means teams are only allowed to bring in new players during a certain time of year. The Premier League also functions on a system very similar to other significant leagues across Europe, such as LaLiga, Bundesliga, Ligue 1 and Serie A.

Benefits and Detriments of League Structure

Like the NHL, there are both pros and cons to the system the premier league has. For the pros, the biggest one is that anyone can own a team. This creates both community support around local teams and much more passionate fans then in North America, and also prevents just the rich few from having a spot in the sport. Another pro is that teams are forced to try to be

successful, instead of situations like in the MLB where the owners of the Colorado Rockies do not care about the team. Another advantage of relegation is that fans are given a reason to go to games when the team is not doing well. There are however also problems with how the EPL operates.

One is that teams can be bought by anyone, leading to significant gaps between clubs who are backed by extremely rich owners and those who aren't. Both Newcastle United and Manchester City are owned by middle eastern oil tycoons, and this increase in funding has allowed them to be major players on the stage of the EPL solely due to money. Another thing is because of no draft and that teams are relegated at the end, the FA system creates dynasties as opposed to a team being able to win the league shortly after finishing poorly. Manchester United ruled the 2000s whereas Manchester City ruled the 2010s due to the fact that it is a lot harder for a new team to become relevant. Another con talent who is premier league quality does not necessarily end up in the league, either because of foreign teams or that player being relegated and not transferring towards a new team. Lastly, the winner being decided by the winner of the league can ruin excitement, as in some years everyone will know the winner of the league 10 games before the season is over.

ROSHN Saudi League

History of the League

The ROSHN Saudi League is the top tier of soccer/football in Saudi Arabia. It was founded in 1976, and government control in Saudi Arabia gives limited information on the league. The Saudi Arabian Football Federation (SAFF) however was founded in 1956. Its history is much more lacking compared to the NHL and EPL. Famous players such as Cristiano Ronaldo, Karim Benzema, and Ngolo Kante currently play in the league, while Neymar played there previously. What makes the Saudi League special as compared to other soccer leagues across the world is that it is funded by the government to help diversify the Saudi economy, which is very reliant on oil. 4 teams receive a lot of significant money, those being Al Ahli, Al Hilal, Al Nassr and Al Ittihad. To promote tourism, the league focuses on signing big names to bring people. Their first major acquisition was Ronaldo, who is arguably one of the five most famous people in the entire world. After that, tons of players continued to come over. These players were enticed by extremely high wages that the league can offer due to their oil money. This has worked in making the quality of the league decent, as in the most recent Club World Cup Al Hilal drew with Real Madrid and beat Manchester City, two of the elites in European Soccer. However, because of how the funding is split up, there is a massive gap between the top four teams and everyone else, which holds the league back from competing with the top leagues in Europe

Organization and Structure

The Saudi League generally operates on a very similar system to the Premier League, with promotion/relegation, transfers, and an open league system and no CBA. The two significant differences between the two is the Saudi League does not have any Financial Fairplay rules, which allows them to sign as many big names as they want unlike the premier league, and the four teams getting government funding to buy players. It is almost a “fake league”, as it is not there for the sport, but to diversify the economy. Money is generated from broadcasting, sponsorships, matchday revenue and commercial ventures like merchandising.

Benefits and Detriments of League Structure

There are both pros and cons to the Saudi League. Firstly, it shares similarities to the Premier league in both good and bad. There are a few differences however. The biggest one is the lack of financial fairplay rules is a positive for the growth of the sport in the country and world, however removes competitiveness as the teams with the richest owners will do the best. There are also only four tiers, so owning a team is much harder and it is not a true open league the way the EPL is. The government funding towards the four main teams also removes competitiveness, as those four teams are naturally going to be better. The idea of selling does also exist, which does allow smaller clubs to be competitive, but not the same way they are in the EPL. The fact that the league isn't created out of love for the sport and instead for the economy means that their fanbase also is not nearly as good as the EPL. The big names do however generate a lot of money, as regardless of who they play for, players like Ronaldo will always generate revenue.

College Football

History of and About NCAA & College Football Playoff

NCAA (National Collegiate Amateur Association) College Football is very different from any other leagues previously mentioned. To start off, the first game was in 1869 between Rutgers and Princeton, but the NCAA, the governing body of college sports, was not founded until 1906. The NCAA was founded to establish rules for college sports and to protect young athletes. All players in the NCAA are considered amateurs and not professionals, as they are viewed as students. Currently, 12 teams make the playoffs, with 4 automatic bids to the 4 highest rated conference championship winners, and the 12th seed given to the 5th highest rated conference winner. There are two major trophies in the sport, those being the national championship and your respective conference championship.

League Structure and Organization

There are 4 major conferences, often called the Power 4, who dominate college football. Those are the ACC (Atlantic Coast Conference), the Big 10, the Big 12, and the SEC. There used to be 5 power conferences, but the Pac-12 fell apart and their major teams joined other conferences. There is no CBA or Salary Cap in college football, and players are considered

amateurs as opposed to employees. There are different divisions, and teams can change divisions, but it is not in the same way the EPL and Roshn Saudi League have division changes via promotion and relegation. Divisions change here because you submit a claim to the NCAA to change either lower or higher. Division 1 FBS is the top division, with Division 1 FCS following, and then Division 2 and Division 3. These levels can play against each other if they schedule each other for out of conference games. There is reward money given to the conference of the winning team for the national championship of FBS. Some schools, like Notre Dame, do not have a conference. The revenue streams differ a lot from school to school. Generally, ticket sales, sponsorships, and merchandise provide schools with money, whereas for conferences TV contracts are their primary sources. A new thing in the sport is NIL money, which allows players to be paid for playing football. This has drastically changed the sport, and teams like Oregon, Ohio State, and Michigan and other big 10 teams with very wide sports programs have become about equal to the SEC (Southeastern conference), who used to dominate college football. Scholarships are another way to entice players to play.

Benefits and Detriments of Structure

College football is truly a unique case, as there is not anything similar. Some of the benefits include giving players a degree and second path in case the sport does not work out. Another benefit is that it creates great community and fanbases as people gain connections to a school and root for them. It is also a great developmental league, as it is a great step between high school and professional football. The biggest detriment is since players are amateurs and not professionals, they are not allowed to have anything like a CBA which can enforce restrictions. This means that the teams with the most funding are now able to spend as much as they want in the era of NIL. It also suffers from being lower in quality compared to the NFL. Lastly, players are sometimes not able to pay due to academic probation, which means that players are bound to another stress outside of the sport.

Comparing Leagues

Now the differences between leagues and role models need to be understood. The most “grassroots” league is the EPL, as they primarily develop inhouse and it starts from a young age with academies. The polar opposite of that is the NHL, which does not have youth academies directly tied to clubs and players are acquired as opposed to grown. The Saudi League is in the middle, as it both develops a significant number of players, especially at lower tiers, but it does gain most of its fame from players who are bought and not developed. College football is a peculiar case, as if offered players can choose any school they want. However, a lot of players do end up going towards their in-state schools, and the scouting is best for teams of players in state out of high school.

Economically, the EFL and the Saudi League are very capitalistic, as teams are rewarded based off success, and punished for doing bad, whereas the American system represents something more similar to communism and socialism, where the want for parity causes talent to

be distributed fairly evenly, with things like the draft lottery. The size and funding causes huge differences in player pay too- generally in the order of EPL, NHL, Saudi League, College football, with the exception being the superstars that get insane amounts of money from Saudi Arabia. These comparisons are essential to understanding the differences between watching different leagues. Different leagues work differently, with pros and cons to each.

Conclusion

Many watch these sports, including multiple of them, but don't understand the fundamental economics that keep the leagues afloat. Different systems vary significantly to each other, and these comparisons make it obvious. In summary, while people might be a fan- few understand the basics of these leagues, and that should be changed.

Limitations

With more time, this research could have more stats to back it up. The writing of this was limited by a short time span, about 12 weeks. This meant that all data was sourced from outside sources, instead of being gathered. The biggest one would be statistics on fan engagement and attendance, which some leagues do not have published information. There are also so many more leagues that deserve attention, and only using four as a baseline for comparison does not allow fair comparisons for some leagues that might not be represented at all here.

Directions for Future Research

To really expand on this paper, the stats like mentioned above would help dig into numbers. Other things like more specific revenue sources would also help. This paper answers a lot of basic questions, however, it barely scratches the surface on economic differences between leagues across the globe.

-

Works Cited

- NHL. (2024). A Brief History of the NHL. www.nhl.com.
<https://www.nhl.com/history/a-brief-history-of-the-league>
- Rivera, J. (2025, June 4). Stanley Cup winners by team: Who has the most championships in NHL history? [Sportingnews.com](https://www.sportingnews.com/us/nhl/news/stanley-cup-winners-team-most-championships-nhl-history/gdv5rchramyobabdxby55x).
<https://www.sportingnews.com/us/nhl/news/stanley-cup-winners-team-most-championships-nhl-history/gdv5rchramyobabdxby55x>
- NHLPa. (2013). Collective Bargaining Agreement. [NHLPa.com](https://www.nhlpa.com/the-pa/cba).
<https://www.nhlpa.com/the-pa/cba>
- Britannica. (2019). Premier League | History & Formation | Britannica. In *Encyclopædia Britannica*. <https://www.britannica.com/topic/Premier-League>
- The Football Association. (2024). The History of The FA. [Www.thefa.com](http://www.thefa.com).
<https://www.thefa.com/about-football-association/who-we-are/history>
- UEFA Champions League - winners and history. (2018). [Footballhistory.org](https://www.footballhistory.org/tournament/champions-league.html).
<https://www.footballhistory.org/tournament/champions-league.html>
- European Cup - winners and history. (2025). [Footballhistory.org](https://www.footballhistory.org/tournament/european-cup.html).
<https://www.footballhistory.org/tournament/european-cup.html>
- Sritharan, J. (2023, June 10). How many English clubs have won the treble? List of teams to win three trophies in a season. [Sportingnews.com](https://www.sportingnews.com/us/soccer/news/how-many-english-clubs-won-treble-list-teams-three-trophies/ro0qh6e0izfhcj1prieoyk4).
<https://www.sportingnews.com/us/soccer/news/how-many-english-clubs-won-treble-list-teams-three-trophies/ro0qh6e0izfhcj1prieoyk4>
- Statista. (2019). Premier League winners/champions 1889-2019 | Statista. Statista; Statista.
<https://www.statista.com/statistics/383696/premier-league-wins-by-team/>
- FA Cup winners/champions 2022. (n.d.). Statista.
<https://www.statista.com/statistics/383679/fa-cup-wins-by-team/>
- UEFA.com. (2023). Most Titles | History | UEFA Champions League. [UEFA.com](https://www.uefa.com/uefachampionsleague/history/winners/).
<https://www.uefa.com/uefachampionsleague/history/winners/>
- The History of the Saudi Pro League: A Comprehensive Overview. (2023, July 31). [Saudifantasyfootball.com](https://saudifantasyfootball.com/blog/the-history-of-the-saudi-pro-league-a-comprehensive-overview).
<https://saudifantasyfootball.com/blog/the-history-of-the-saudi-pro-league-a-comprehensive-overview>
- How the Saudi Pro League Generates Income and Attracts Top Football Talent. (2025, April 5). [Notordinaryblogger.com](https://notordinaryblogger.com/how-the-saudi-pro-league-generates-income-and-attracts-top-football-talent/); Notordinaryblogger.
<https://notordinaryblogger.com/how-the-saudi-pro-league-generates-income-and-attracts-top-football-talent/>
- Schrag, T., & Content, B. (2025, June 17). The Economics of College Athletics: Understanding Revenue Streams and Expenditures - Kansas State Collegian. [Kansas State Collegian](https://kstatecollegian.com/2025/06/16/the-economics-of-college-athletics-understanding-revenue-streams-and-expenditures/).
<https://kstatecollegian.com/2025/06/16/the-economics-of-college-athletics-understanding-revenue-streams-and-expenditures/>
- Maysoon. (2025, April 16). Football Clubs in Saudi Arabia | A Story of Legacy, Grit, and

Growth. Soul of Saudi. <https://soulofsaudi.com/football-clubs-saudi-arabia/>
NCAA.com. (2024, September 8). How the 12-team College Football Playoff works.
NCAA.com.
<https://www.ncaa.com/news/ncaa/article/2025-01-20/college-football-history-notable-firs>
ts-and-milestones
NCAA. (2021, May 4). History. NCAA.org. <https://www.ncaa.org/sports/2021/5/4/history.aspx>
Must-Know Sport Viewership Statistics [Latest Report] • Gitnux. (n.d.). Gitnux.org.
<https://gitnux.org/sport-viewership-statistics/>

Dorsal Fin Edge Proportions and Their Relationship to Swimming Strategy in Sharks

By Benjamin Janitz

Abstract

Dorsal fins play a critical role in elasmobranch locomotion by providing stability, reducing roll, and enhancing swimming efficiency. Variation in dorsal fin geometry may therefore reflect ecological niche and swimming strategy. This study quantitatively compares dorsal fin leading-to-lagging edge ratios in three shark species with differing ecological roles: the spiny dogfish (*Squalus acanthias*), salmon shark (*Lamna ditropis*), and longfin mako (*Isurus paucus*). Using lateral-view photographs, dorsal fin leading and trailing edges were measured and expressed as proportional ratios to account for differences in overall body size. Results revealed substantial interspecific variation: the spiny dogfish exhibited the most significant leading-edge proportion (137.5%), the salmon shark showed an intermediate ratio (116.67%), and the longfin mako displayed a symmetrical fin geometry (100%). These differences closely correspond to each species' swimming behavior and habitat use. The elongated leading edge of the spiny dogfish enhances maneuverability at low speeds, the salmon shark's moderately extended leading edge supports high-speed pursuit and stabilization, and the longfin mako's symmetrical fin promotes efficient cruising during long-distance migration. These findings demonstrate that dorsal fin morphology is an adaptive trait shaped by hydrodynamic and ecological demands, highlighting the functional relationship between fin geometry and swimming performance in sharks.

Introduction

In the order Elasmobranchii, the dorsal fin, located on the dorsal side of the shark, skate, or ray, serves several purposes, such as stabilizing the organisms from rolling onto their sides and aiding in swimming in a straight-forward motion. The trailing edge of the dorsal fin may also create a low-pressure area that extends to the tail, increasing caudal fin efficiency and forward thrust while conserving energy (State of Hawaii). A comparative assessment of dorsal fin edge proportions between species may reveal how fin geometry reflects ecological niche and swimming mode. Because elasmobranch fins function as hydrofoils that generate lift, resist roll, and reduce energetic cost, variation in the relative lengths of the leading and lagging edges should correlate with maneuverability, cruising efficiency, or burst-speed performance.

This study aims to quantify and compare the dorsal fin leading-to-lagging edge ratios in three shark species—*Squalus acanthias*, *Isurus paucus*, and *Lamna ditropis*—to determine whether fin shape is associated with differences in habitat, swimming behavior, and ecological specialization. And while individual dorsal fin morphology studies have been done, there are no specific published studies that directly and exclusively compare the morphology of the dorsal fins across all three species (spiny dogfish, salmon shark, and longfin mako) in a single, dedicated comparative analysis.

It is predicted that species that rely on maneuverability (e.g., spiny dogfish) will exhibit larger leading-edge proportions, species optimized for rapid acceleration and sustained speed (e.g., salmon sharks) will show moderate but streamlined edge ratios, and species that cruise efficiently over long distances (e.g., longfin mako) will display more symmetrical edge proportions that minimize drag during constant swimming.

Methods and Materials

Using the photographs of the spiny dogfish (*Squalus acanthias*), longfin mako shark (*Isurus paucus*), and the salmon shark (*Lamna ditropis*), the leading and lagging edges of the dorsal fins were measured in centimeters. The measurements begin at the tip of the dorsal fin and extend to the anterior point of the leading edge base, and the lagging edge is measured from the dorsal fin tip to the posterior point of the base. The leading edge is the numerator, and the lagging edge is the denominator; dividing the leading edge by the lagging edge and multiplying the quotient by 100 yields a percent length. The percent length given represents the percentage of the lagging edge that the leading edge measures. Because all dorsal fin measurements were derived from images rather than physical specimens, absolute size values cannot be compared directly between species. For this reason, fin-edge ratios were used instead of raw lengths. Only fins photographed in lateral view with minimal distortion were selected to reduce parallax error. All measurements were taken three times and averaged to minimize observer variability.

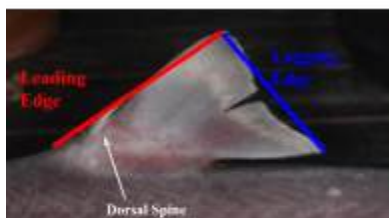
Figure 1

Common Name: Spiny Dogfish

Scientific Name: *Squalus acanthias*

Leading Edge Length: 5.5 units

Lagging Edge Length: 4.0 units



Shown above is the first dorsal fin along with the first dorsal spine of a spiny dogfish (Pfleger)

Figure 2

Common Name: Salmon Shark

Scientific Name: *Lamna ditropis*

Leading Edge Length: 3.5 cm

Lagging Edge Length: 3.0 cm



Shown above is the first dorsal fin of a salmon shark (Olson).

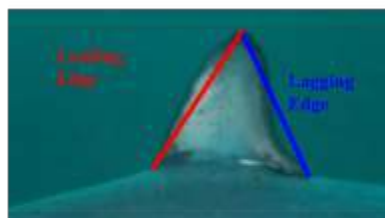
Figure 3

Common Name: Longfin Mako Shark

Scientific Name: *Isurus paucus*

Leading Edge Length: 4.0 cm

Lagging Edge Length: 4.0 cm



Shown above is the first dorsal fin of a longfin mako shark (Oceana).

Results

The dorsal fin of the Spiny Dogfish appears as an obtuse triangle made up of thin cartilage; it protrudes from the dorsal side of the shark with a spine that lifts slightly from the dorsal fin but still follows the angle of the anterior edge. The fin is light brown at the base and somewhat opaque as it becomes further from its origin along the shark's back; this dorsal fin also has small lines that follow the angle of the spiny dogfish dorsal fin, seemingly with a slight

corrugated texture and moderate stiffness when swimming (Maia and Wilga). The dorsal fin of the Salmon shark appears to be closer to an acute triangle and is made of thicker cartilage than the Spiny Dogfish. The Salmon Shark's dorsal seems to be a much darker brown—nearly black along the outside of the fin with a speckled gradient to a lighter brown towards the center of the fin (Abercrombie et al.). The dorsal fin of the longfin mako shark resembles an equilateral triangle, with three acute angles. The coloration is a grey-teal, with a solid dark teal pigmentation along the outside of the fin, quickly grading into a light grey or blue towards the center. Similar to the dorsal fin of the salmon shark, the dorsal fin of the longfin mako shark is relatively thick and smooth (Abercrombie et al.).

The dorsal fin of the Spiny Dogshark measured roughly 5.5 units at the leading edge and 4 units at the lagging edge (Refer to Figure 1), yielding a ratio of 5.5:4 and a quotient of 1.375. The quotient multiplied by 100 gives the percent length of the leading edge to the lagging edge. For the Spiny Dogfish, the leading edge is 137.5% of the lagging edge, meaning that the leading edge is 37.5% longer than the lagging edge of the dorsal fin. The leading edge of the Salmon Shark dorsal fin measured 3.5 units, with the lagging edge measuring 3.0 units (Refer to Figure 2); this gave a ratio of 3.5:3.0, or 1.1667 when the leading edge is divided by the lagging edge. The quotient is then multiplied by 100 to yield 116.67%; this percentage tells that the leading edge of the dorsal fin is 116.67% of the length of the lagging edge, meaning that the leading edge is 16.67% longer than the lagging edge. The dorsal of the Mako Shark measured 4.0 units on the leading edge and 4.0 units on the lagging edge (Refer to Figure 3); this gives a 4.0:4.0 ratio or 1.0:1.0. This equates to the length of the leading edge to be 100% of the lagging edge, meaning that the two are the same length from the points measured.

Across the three species, dorsal fin edge proportions varied considerably. The spiny dogfish possessed the most significant proportional leading edge (137.5%), followed by the salmon shark (116.67%), while the longfin mako exhibited a perfectly symmetrical fin (100%). These values indicate a progression from highly maneuverability-oriented fin geometry (dogfish), to mixed-performance geometry (salmon shark), to streamlined, cruising-adapted geometry (longfin mako).

Discussion

The spiny dogfish has two moderately sized dorsal fins, each with a sharp spine that is used for protection. The dorsal fins provide stability during moderate-speed swimming, while the spine serves as a deterrent to predators (Bester). Spiny dogfish are benthic species that inhabit deep coastal waters. The dogfish swims in slow-moving schools; not a fast swimmer, it relies on its fins for maneuverability rather than speed. The fin size helps the spiny dogfish maintain balance and control of the variable water conditions it inhabits globally. The small and flexible dorsal fin of the dogfish, with the leading edge 137.5% of the lagging edge, allows the shark to have increased maneuverability at slow speeds.

The salmon shark is found in cold, open-ocean waters and coastal waters throughout the North Pacific Ocean. It is commonly found near Japan, North Korea, and Russia, as well as

along the Bering Sea (Roman). Being primarily a pelagic species, the salmon shark is usually found near the surface or at depths of around 500 ft (152 m), and has been photographed at depths of up to 837 ft (255 m) (Roman). Like the spiny dogfish, the salmon shark is a migratory species; however, its migration patterns tend to align with those of other pelagic fish it preys upon, such as Pacific salmon. The salmon shark's dorsal fin is large, upright, and pointed; this helps the shark maintain a streamlined body and swim at high speeds to catch prey. The ability to swim fast is needed because the sharks' prey, the Pacific salmon, tends to elude the shark in high-speed chases; these chases also involve rapid movement, which the dorsal fin counteracts to aid the shark's stabilization (Roman). The salmon shark's leading edge of its dorsal fin, which is 116.67% the length of its lagging edge, helps build a strong, upright fin that supports its lifestyle. Inhabiting deep, offshore tropical and subtropical waters, the longfin mako shark tends to cruise slowly rather than at high speeds. The mako tends to have a relatively small dorsal fin, reflecting its energy-efficient movement as an open-ocean fish. The dorsal fin of the longfin mako aids more in the shark's directional control than stability. The longfin mako is considered highly migratory, but little is known about them due to their tendency to be misidentified as their cousin, the shortfin mako. Being highly migratory and an open-ocean fish, the longfin mako depends on swimming long distances and cruising; this makes the large dorsal fin, which aids white sharks and salmon sharks, redundant. The smaller fin, instead, aids in the shark's direction. Directional control is aided by the equilateral triangular shape of the makos fin, which helps regulate flow and streamline the shark for continuous swimming (Yagmur et al.). The longfin mako shark's dorsal fin measured 4.0 units on its leading edge and 4.0 units on its lagging edge; this makes a 4.0 to 4.0 ratio or a 1.0:1.0, meaning that the leading edge is 100% the length of the lagging edge or that they are the same length. This evenly distributed dorsal fin aids in the longfin mako shark's constant movement through the open ocean and its long migrations.

When viewed comparatively, the relationship between dorsal fin geometry and ecological strategy becomes clear. The spiny dogfish, a slower-moving benthopelagic shark, benefits from an elongated leading edge because it increases the surface area that initiates flow attachment. This enables enhanced control at low speeds and reduces the likelihood of roll when maneuvering in heterogeneous coastal environments. In contrast, the salmon shark inhabits high-energy pelagic waters and frequently lunges at salmon; the moderately extended leading edge (116.67%) provides sufficient surface area to stabilize high-speed turns while maintaining a narrow posterior edge to reduce drag. The longfin mako's symmetrical 1:1 ratio reflects a hydrodynamic compromise optimized for steady cruising rather than burst performance. A fin with equal leading and trailing edges promotes consistent flow attachment along both surfaces, reducing induced drag during long-distance migrations. These patterns demonstrate that dorsal fin morphology among elasmobranchs is not arbitrary but closely tied to ecological niche, swimming mechanics, and energetic demands.

Conclusion

This study demonstrates a clear association between dorsal fin edge proportions and the ecological and behavioral demands of three shark species. The spiny dogfish exhibits a long leading edge suited for maneuverability and stabilization at slow speeds. The salmon shark's moderately extended leading edge reflects its need for stability during rapid, high-energy pursuit. The longfin mako's symmetrical fin supports efficient cruising during long migrations. These findings emphasize that dorsal fin morphology is an adaptive trait that reflects the hydrodynamic challenges posed by each species' environment. Further research with larger sample sizes and three-dimensional fin scans would improve precision and deepen understanding of fin function in elasmobranch locomotion.

Acknowledgments

I would like to thank Dr. Ashley Stoehr for her guidance and support throughout the design and execution of this study. Appreciation is extended to the Shoals Marine Laboratory for inspiring this research through their focus on aquatic ecosystems and field-based inquiry.

Works Cited

- Abercrombie, Debra L., et al. "Visual Identification of Fins from Common Elasmobranchs in the Northwest Atlantic Ocean." NOAA, Mar. 2013, media.fisheries.noaa.gov/dam-migration/sharkfinid_guide.pdf. Accessed 7 Aug. 2025.
- Bester, Cathleen. "Spiny Dogfish." Florida Museum of Natural History, 12 Feb. 2025, www.floridamuseum.ufl.edu/discover-fish/species-profiles/spiny-dogfish/. Accessed 8 Aug. 2025.
- Maia, Anabela, and Cheryl D. Wilga. "Anatomy and Muscle Activity of the Dorsal Fins in Bamboo Sharks and Spiny Dogfish during Turning Maneuvers." *Journal of Morphology*, vol. 274, no. 11, 2 Aug. 2013, pp. 1288-98, <https://doi.org/10.1002/jmor.20179>. Accessed 7 Aug. 2025.
- Oceana. "Longfin Mako Shark." Oceana, oceana.org/marine-life/longfin-mako-shark/. Accessed 7 Aug. 2025.
- Olson, Brooke. "Salmon Shark." Washington Department of Fish & Wildlife, 2025, wdfw.wa.gov/species-habitats/species/lamna-ditropis. Accessed 7 Aug. 2025.
- Pfleger, Mariah. "Shark Fact Friday – A Dogfish Named Spot." Oceana USA, 23 Feb. 2018, usa.oceana.org/blog/shark-fact-friday-dogfish-named-spot/. Accessed 7 Aug. 2025.
- Roman, Brenda. "Salmon Shark." Florida Museum of Natural History, 5 Feb. 2025, www.floridamuseum.ufl.edu/discover-fish/species-profiles/salmon-shark/. Accessed 8 Aug. 2025.
- State of Hawaii. "Fins & Swimming." Hawai'i Sharks, Hawai'i Department of Land and Natural Resources, dlnr.hawaii.gov/sharks/anatomy/fins-swimming/. Accessed 7 Aug. 2025.
- Yagmur, Sercan, et al. "Comparison of Flow Characteristics around an Equilateral Triangular Cylinder via PIV and Large Eddy Simulation Methods." *Flow Measurement and Instrumentation*, vol. 55, June 2017, pp. 23-36, <https://doi.org/10.1016/j.flowmeasinst.2017.04.001>.

Predicting ecDNA Presence with Machine Learning on Inexpensive Genomic Data

By Esha Verma¹, Azhar Khandekar²

Abstract

Extrachromosomal DNA (ecDNA) is an important mechanism of oncogene amplification in cancer and is associated with increased tumor aggressiveness, genetic heterogeneity, and poor patient outcomes. While ecDNA has traditionally been studied in advanced cancers, recent evidence suggests that it may also arise in precancerous cells, increasing interest in early and accessible detection methods. Current techniques for identifying ecDNA, including whole-genome sequencing (WGS) and fluorescence in situ hybridization (FISH), are accurate but costly and not widely available. In this study, we performed a feasibility analysis to evaluate whether ecDNA presence can be predicted using low-cost SNP 6 array-derived copy number data and machine learning. We formulated ecDNA detection as a segment-level binary classification problem using features derived from copy number segments, including LogR, B-allele frequency (BAF), total copy number, segment size, and differences in copy number and BAF between adjacent genomic segments. Segment-level ecDNA labels were obtained from AmpliconArchitect predictions based on matched WGS data and were treated as proxy ground truth. We trained and evaluated logistic regression, random forest, and neural network classifiers using five-fold cross-validation. Across all models, no ecDNA-positive segments were correctly identified, resulting in zero sensitivity for the ecDNA class and a mean accuracy of approximately 0.49, consistent with random classification. Although some models achieved moderate area-under-the-ROC-curve (AUROC) values, further analysis revealed that this performance was driven almost entirely by correct classification of non-ecDNA segments. These results demonstrate that copy number features derived from SNP 6 arrays alone do not provide sufficient information to reliably distinguish ecDNA from other copy number alterations. Our findings highlight fundamental limitations of low-resolution genomic data for ecDNA detection and suggest that future approaches will require higher-resolution structural information or integration of additional data types.

1. Introduction

1.1 Background on Extrachromosomal DNA

Extrachromosomal DNA (ecDNA) consists of circular DNA fragments that exist outside the linear chromosomal genome. These structures often carry oncogenes and can occur at high copy number, enabling rapid oncogene amplification and increased gene expression. Unlike chromosomal DNA, ecDNA lacks centromeres and is therefore unevenly inherited during cell division, which contributes to intratumoral heterogeneity and rapid tumor evolution. Previous

¹ Saratoga High School, Saratoga CA

² Hutchinson Cancer Center, Seattle Washington

studies have shown that ecDNA is prevalent across many cancer types and is strongly associated with poor patient outcomes and therapeutic resistance. Historically, ecDNA was believed to arise primarily in late-stage tumors. However, recent evidence suggests that ecDNA can also form in precancerous or early-stage lesions. This observation raises the possibility that earlier detection of ecDNA could have clinical importance. As a result, there is growing interest in developing scalable and cost-effective computational methods for ecDNA identification.

1.2 Current Approaches to ecDNA Detection

Most established approaches for ecDNA detection rely on whole-genome sequencing (WGS) combined with specialized computational tools, such as AmpliconArchitect, or on laboratory-based imaging techniques including fluorescence in situ hybridization (FISH) and microscopy. While these methods provide high accuracy and detailed structural information, they are expensive, resource-intensive, and not routinely accessible in many clinical or research settings. In contrast, SNP 6 copy number arrays are widely used in clinical genomics due to their relatively low cost and broad availability. These arrays capture genome-wide copy number alterations using probe intensity (LogR) and allelic imbalance measurements such as B-allele frequency (BAF). Because of their accessibility, SNP arrays represent a potential platform for large-scale screening. However, it remains unclear whether features derived from SNP array data contain sufficient information to reliably distinguish ecDNA from chromosomal copy number amplifications.

1.3 Study Objective

The objective of this study was to perform a feasibility analysis to determine whether ecDNA presence can be predicted using only features derived from SNP 6 copy number array data. We evaluated the performance of multiple machine learning models on a segment-level binary classification task and examined the impact of low-resolution genomic features and severe class imbalance on model performance.

2. Materials and Methods

2.1 Dataset Description

We analyzed somatic copy number alteration data derived from lung cancer tumor samples for which matched whole-genome sequencing (WGS) data were available. The dataset consisted of 20,812 copy number segments obtained from 173 unique tumor samples. Each copy number segment was defined by its chromosome number (restricted to autosomes), genomic start coordinate, and genomic end coordinate. Each row in the dataset corresponded to a single segmented region of uniform copy number across the genome. Segment-level ecDNA labels were obtained using AmpliconArchitect, a computational tool that reconstructs the architecture of amplified genomic regions from WGS data. Copy number segments that overlapped regions predicted to be ecDNA by AmpliconArchitect were labeled as ecDNA-positive, while all other

segments were labeled as non-ecDNA. These labels were treated as proxy ground truth for model training and evaluation. Because ecDNA status was inferred computationally rather than experimentally validated for every segment, the labels may contain noise resulting from prediction uncertainty or incomplete structural resolution.

2.2 Feature Engineering

All features used for model training were derived exclusively from SNP 6 array copy number data. The initial feature set included probe intensity (LogR), which reflects relative DNA copy number, B-allele frequency (BAF), which captures allelic imbalance, total copy number (TCN), and the genomic size of each copy number segment. These features are commonly used to characterize copy number alterations in cancer genomes. To incorporate information about local genomic transitions, we engineered additional features based on differences between adjacent copy number segments. Specifically, we calculated the difference in total copy number between each segment and the immediately preceding segment along the genome. This feature was designed to capture abrupt increases or decreases in copy number that may reflect structural discontinuities. In addition, we computed the difference in B-allele frequency between adjacent segments using the same sequential approach. Sudden shifts in BAF may reflect changes in allelic balance that accompany complex genomic rearrangements. These difference-based features were intended to provide contextual information beyond absolute copy number values by allowing the model to recognize patterns across neighboring genomic regions. All engineered features were computed at the segment level and used as inputs to the machine learning classifiers.

2.3 Class Imbalance Handling

In real-world biomedical datasets, particularly in disease detection tasks, class imbalance is a common challenge due to the relatively low prevalence of positive cases compared to negative cases. In such settings, machine learning models may become biased toward the majority class, resulting in high apparent accuracy but poor performance in identifying the minority class, which is often of greater scientific or clinical interest. In our dataset, ecDNA-positive segments represented a very small fraction of the total data. Specifically, 232 copy number segments were labeled as ecDNA-positive, whereas 19,676 segments were labeled as non-ecDNA. During initial model training, this severe class imbalance caused the models to favor predictions of the non-ecDNA class, leading to high accuracy for non-ecDNA segments but poor detection of ecDNA-positive segments. To address this imbalance, we evaluated different class-balancing strategies. Our primary approach was random oversampling of the ecDNA-positive class in the training data. Random oversampling increases the representation of the minority class by duplicating existing ecDNA-positive samples until a more balanced class distribution is achieved. This approach preserves all original training samples while reducing bias toward the majority class. We also evaluated random downsampling of the non-ecDNA class, in which the majority class is reduced to match the size of the minority class. However,

because this approach required discarding a large number of non-ecDNA samples, it reduced the effective size of the training dataset by approximately two orders of magnitude. As a result, downsampling led to decreased model stability and poorer overall performance compared to oversampling. Based on these observations, random oversampling of ecDNA-positive segments was selected as the primary method for addressing class imbalance in this study.

2.4 Machine Learning Models

We evaluated three supervised machine learning models for binary classification of ecDNA presence at the segment level: logistic regression, a neural network, and a random forest classifier. These models were selected to represent a range of model complexities, from linear to non-linear approaches.

Logistic regression was used as a baseline model due to its interpretability and widespread use in biomedical classification tasks. The dataset was split into training and testing subsets using an 80%–20% split. Model performance was evaluated using k-fold cross-validation, as described in Section 2.5. In addition, an independent validation dataset consisting of 60 cell lines with experimentally validated ecDNA status obtained from orthogonal imaging techniques was used to assess generalizability where applicable.

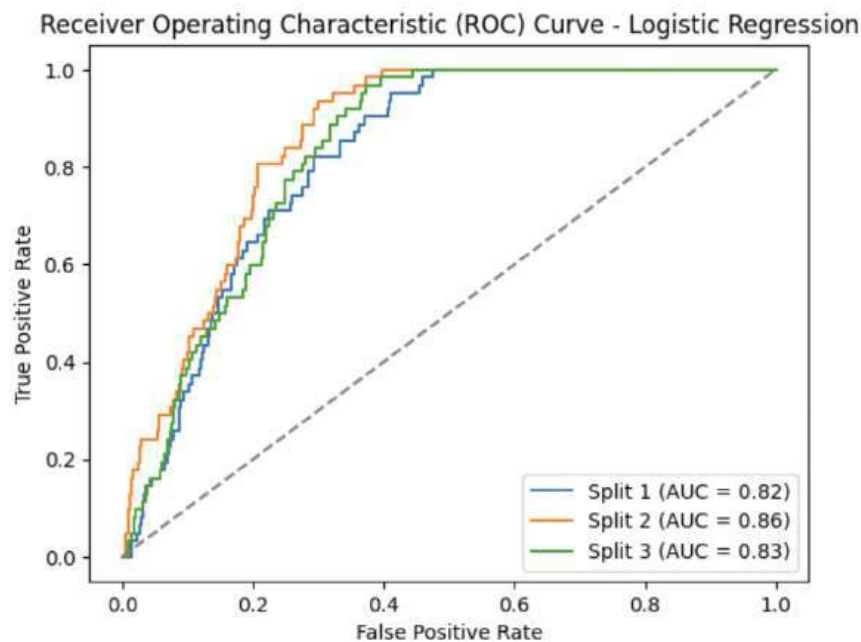


Figure 1. Receiver operating characteristic (ROC) curves for detecting the presence of ecDNA using logistic regression with k-fold cross-validation. ROC curves show true positive rate versus false positive rate across three cross-validation splits. A logistic regression classifier was trained to predict ecDNA presence, and model performance was evaluated using the area under the ROC curve (AUC). The dashed diagonal line represents random classification. AUC values for each split are indicated in the legend (Split 1 = 0.82, Split 2 = 0.86, Split 3 = 0.83).

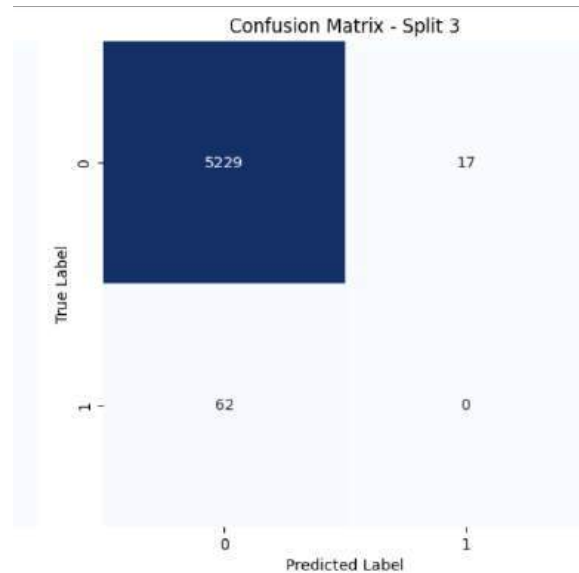


Figure 2: Confusion Matrix for the third split in the logistic regression model.

To model more complex, non-linear relationships between features, we also implemented a neural network classifier. The neural network consisted of fully connected layers with non-linear activation functions. A sigmoid activation function was applied to the output layer to produce probability estimates for binary classification. The model was trained using binary cross-entropy loss, which is commonly used for binary classification tasks and penalizes incorrect predictions more strongly as the confidence of the prediction increases. Model parameters were optimized through iterative training on the training data.

In addition, we evaluated a random forest classifier, which is an ensemble learning method that combines predictions from multiple decision trees. Each decision tree was trained on a bootstrap sample of the training data, generated by sampling with replacement. At each split within a tree, a random subset of features was considered, introducing feature-level randomness that helps reduce overfitting and model bias. For classification, the final prediction of the random forest was determined by a majority vote across all individual trees. Key hyperparameters of the random forest model included the number of trees, the maximum depth of each tree, and the number of features considered at each split.

2.5 Cross-Validation Strategy

We used cross-validation to evaluate the performance and robustness of the machine learning models on unseen data. Cross-validation is a commonly used technique in machine learning that helps reduce bias introduced by a single train-test split and provides a more reliable estimate of model performance. For model training, the dataset was first shuffled and then divided into multiple subsets according to the k-fold cross-validation procedure. In this approach, the dataset is partitioned into k equally sized folds. During each iteration, one fold is used as the validation set while the remaining k - 1 folds are used for training. This process is repeated k times so that each fold serves as the validation set exactly once. Model performance metrics are

then averaged across all folds to estimate overall performance. In this study, we used five-fold cross-validation ($k = 5$). Across the five validation folds, the models achieved a mean classification accuracy of approximately 0.49. This evaluation framework allowed us to assess the consistency of model performance and to identify limitations related to class imbalance and feature resolution.

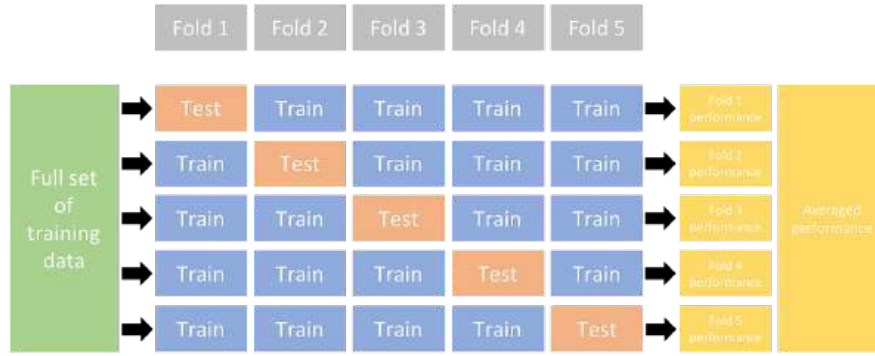


Figure 3: This image explains the process of k - fold cross validation. The full set of training data is split into k numbers and trained. After training of each of these folds, the performance metrics are averaged.

3. Results

Evaluation Metrics

To evaluate the performance of the machine learning models, we analyzed several standard classification metrics, including the confusion matrix, precision, recall, F1 score, and support. Together, these metrics provide a comprehensive assessment of model performance, particularly in the presence of severe class imbalance. The confusion matrix is a 2×2 table that summarizes the number of true positives, true negatives, false positives, and false negatives produced by the classifier. This representation allows for direct examination of how well the model distinguishes between ecDNA-positive and non-ecDNA segments. Based on the confusion matrix, we computed additional evaluation metrics that quantify different aspects of model performance. Across the evaluated models, the confusion matrix revealed 5,229 true negative segments, 62 false negative segments, 0 true positive segments, and 17 false positive segments. The absence of any true positive predictions indicates that none of the models successfully identified ecDNA-positive segments. Although the number of false positives was relatively small, the complete failure to detect ecDNA-positive segments highlights a major limitation of the classification approach when applied to the available features. Precision measures the proportion of predicted positive cases that are truly positive, while recall (also referred to as sensitivity) measures the ability of the model to correctly identify all positive instances. The F1 score represents the harmonic mean of precision and recall and provides a single metric that balances these two measures. Support refers to the number of true samples belonging to each class in the dataset. In this study, the support values differed substantially between classes, reflecting the strong class imbalance present in the data.

	Precision	Recall	F1-score	support
Non-ecDNA	0.99	1.00	0.99	5246
ecDNA	0.00	0.00	0.00	62

As shown in Table 1, the evaluation metrics for the non-ecDNA class were consistently high, with precision, recall, and F1 scores all close to 1.00. In contrast, the ecDNA class exhibited precision, recall, and F1 scores of 0.00, reflecting the model's inability to correctly classify any ecDNA-positive segments. These results indicate that while the models performed well in identifying non-ecDNA segments, they were unable to learn discriminative patterns associated with ecDNA using the available feature set.

Problem of Sensitivity and Specificity

Initial evaluation using receiver operating characteristic (ROC) curves suggested promising model performance, with area under the ROC curve (AUC) values of 0.84 for logistic regression, 0.97 for the random forest model, and 0.64 for the neural network. At first glance, these values appeared to indicate effective discrimination between ecDNA and non-ecDNA segments. However, further analysis revealed that these high AUC values were driven almost entirely by correct classification of the majority class. This observation highlighted the importance of examining sensitivity and specificity in addition to aggregate performance metrics. Sensitivity refers to the ability of a model to correctly identify ecDNA-positive segments, whereas specificity refers to the ability to correctly identify non-ecDNA segments. In this study, the models achieved high specificity but zero sensitivity. Because the classifiers consistently predicted the non-ecDNA class, the number of true negatives was high, which artificially inflated performance metrics such as accuracy and AUC. However, the lack of true positive predictions resulted in complete failure to detect ecDNA. This discrepancy illustrates how conventional performance metrics can be misleading in the presence of severe class imbalance and emphasizes the importance of examining class-specific metrics when evaluating models for rare-event detection.

4. Discussion

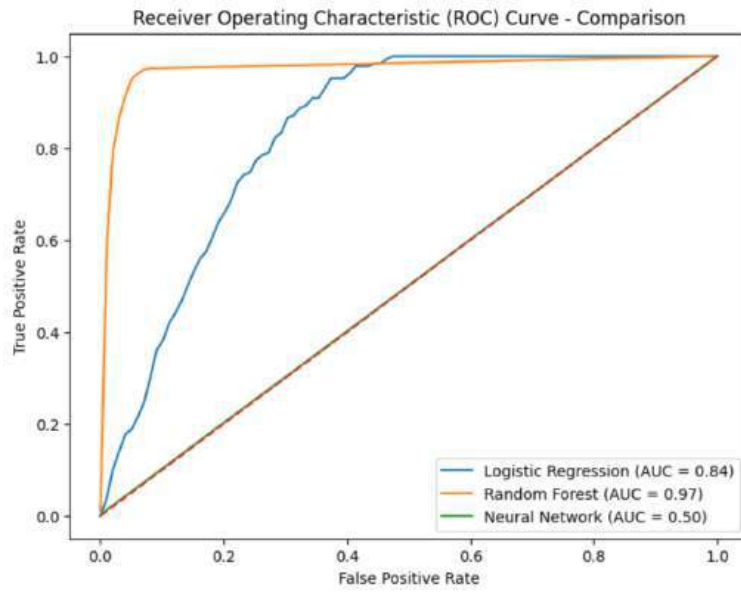


Figure 4: Performance metric of all three machine learning algorithms with k - fold cross validation. Logistic Regression is outlined in blue, random forest in orange, and neural network in green.

In this study, we implemented and evaluated three supervised machine learning models, logistic regression, random forest, and a neural network, with the goal of predicting the presence of extrachromosomal DNA (ecDNA) using data derived from SNP 6 arrays. Among the three models, the random forest classifier produced the highest area under the receiver operating characteristic curve (AUROC). However, upon closer examination, we found that this performance was misleading. The model's predictions overwhelmingly favored the majority class, non-ecDNA segments, resulting in zero detection of ecDNA-positive segments. Consequently, despite superficially high AUROC scores, the models failed to provide meaningful predictions for the minority class.

These results highlight the inherent challenges of detecting ecDNA computationally using low-resolution SNP array data. The binary classification task is complicated by the scarcity of ecDNA-positive segments, the similarity between ecDNA-associated copy number alterations and other genomic amplifications, and the limited information provided by standard SNP array features such as LogR, B-allele frequency (BAF), total copy number (TCN), segment size, and differences in LogR or BAF between consecutive segments. In this context, the models' inability to correctly classify any ecDNA-positive segments suggests that these features alone do not contain sufficient discriminative information to identify ecDNA reliably.

Our findings indicate that more sophisticated feature engineering is essential for improving predictive performance. For example, integrating additional structural or sequence-level features, leveraging multi-omics datasets, or incorporating features derived from complementary experimental techniques (e.g., FISH or whole-genome sequencing) may increase the models' ability to detect ecDNA. Furthermore, the use of advanced machine learning algorithms capable of capturing highly non-linear relationships, such as gradient boosting

frameworks or deep neural networks with specialized architectures, represents a promising avenue for future investigation.

Another critical consideration is the availability and quality of labeled training data. In this study, ecDNA labels were obtained from AmpliconArchitect predictions based on whole-genome sequencing data, which, while informative, are only proxy ground truth and may contain noise. The development and curation of larger datasets with highly accurate, experimentally validated ecDNA annotations will be essential for training models that generalize effectively across diverse cancer types and patient samples.

Overall, our results underscore a fundamental limitation: standard SNP 6 array features are insufficient for reliably predicting ecDNA presence. While machine learning remains a powerful tool for genomic classification tasks, its success depends on both the informativeness of the input features and the availability of robust, well-labeled training datasets. Future studies should focus on integrating richer genomic features and more sophisticated computational approaches to improve ecDNA detection, particularly in settings where low-cost assays like SNP arrays are the only available data source.

5. Conclusion

In this study, we investigated whether ecDNA could be detected using machine learning models applied to SNP 6 copy number array data. Despite testing multiple models and engineered features, including LogR, B-allele frequency, total copy number, segment size, and their differences between consecutive segments, our models were unable to reliably identify ecDNA-positive segments. The low sensitivity of all models highlights a fundamental limitation of using low-resolution SNP array data for ecDNA detection. These findings suggest that the subtle genomic patterns associated with ecDNA require more informative data, such as high-resolution sequencing, structural variant information, or multimodal datasets. Additionally, future work would benefit from larger datasets with experimentally validated ecDNA labels to improve model training and evaluation. Overall, while our approach did not achieve accurate predictions, it demonstrates the importance of feature selection, data quality, and class balance in applying machine learning to genomic research. This study provides a foundation for developing more sophisticated computational strategies to detect ecDNA in cancer and underscores the challenges of using accessible, low-cost genomic data for this purpose.

Works Cited

- Conger, K. (2023, April 10). Tiny DNA circles are key drivers of cancer, Stanford Medicine-led international study finds. Stanford Medicine News Center.
<https://med.stanford.edu/news/all-news/2023/04/ecDNA-cancer.html>
- National Cancer Institute. (2022, April 20). ecDNAs are extra active and terrifying. Cancer Currents Blog.
<https://www.cancer.gov/ccg/blog/2022/interview-ecdna>
- Kim, H., Nguyen, N.-P., Turner, K., Wu, S., Gujar, A. D., Luebeck, J., Liu, J., et al. (2020). Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nature Genetics*, 52(9), 891–897.
<https://doi.org/10.1038/s41588-020-0678-2>
- Luo, J., Li, Y., Zhang, T., Xv, T., Chen, C., Li, M., Qiu, Q., Song, Y., & Wan, S. (2022). Extrachromosomal circular DNA in cancer drug resistance and its potential clinical implications. *Frontiers in Oncology*, 12, 1092705.
<https://doi.org/10.3389/fonc.2022.1092705>
- Verhaak, R. G. W., Bafna, V., & Mischel, P. S. (2019). Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nature Reviews Cancer*, 19(5), 283–288.
<https://doi.org/10.1038/s41568-019-0122-2>
- Zhu, Y., Gujar, A. D., Wong, C.-H., Tjong, H., Ngan, C. Y., Gong, L., Chen, Y.-A., et al. (2021). Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. *Cancer Cell*, 39(5), 694–707.e7.
<https://doi.org/10.1016/j.ccell.2021.03.006>
- Brownlee, J. (2022, July 5). Binary classification tutorial with the Keras deep learning library. *Machine Learning Mastery*.
<https://machinelearningmastery.com/binary-classification-tutorial-with-the-keras-deep-learning-library/>
- Google Developers. (n.d.). Classification: ROC curve and AUC. Google Machine Learning Crash Course. Accessed February 5, 2024.
<https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>

Review of the Impacts of Carbon Nanotubes on the Freeze-Thaw Resistance of Concrete

By Rickie Chen

Abstract

Concrete is the most widely used material in the world today, with an annual production of approximately 14 billion cubic meters of concrete. Despite this, concrete remains susceptible to cracking, weathering, and stress caused by both human activities and environmental factors. Researchers around the world have been seeking effective solutions for decades. One notable proposal is the use of nanotechnology. Nanomaterials measure 100 nanometers or less in size and have the potential to increase the strength and resistance to damage of the concrete. One of the most important types of resistance is freeze-thaw (FT) resistance, which is the ability of concrete to withstand the damaging effects of water freezing and expanding within the concrete's pores during repeated thermal cycles. Although many advancements have already been made, the effect of nanotechnology, such as Multi-walled carbon nanotubes (MWCNT), on the FT resistance of concrete remains unestablished. The goal of this research paper is to explore the role of MWCNTs in concrete's resistance to FT cycles. Through a combination of experimentation and a review of existing scientific literature, this research aims to identify nanotechnological solutions to address the challenges associated with concrete.

Introduction

Concrete has been one of the most widely used construction materials in the world. Its earliest known form dates back to approximately 1300 B.C., when it was created from crude elements combined into a mixture. Over time, scientists and engineers have developed methods to enhance the technological and mechanical properties of concrete by incorporating specific additives, resulting in a more durable and resilient material. Its strength and durability are so significant that, according to Arizona State University, approximately 30 billion tons of concrete are used each year globally in homes, buildings, bridges, and other structures. As such, concrete remains the cornerstone of modern construction (Kullman, 2023).

Despite its widespread utility, concrete has many vulnerabilities. It is susceptible to issues such as cracking, environmental stress, and damage from human activities, all of which can lead to damage to the concrete. Researchers have long sought methods to mitigate these problems and extend the longevity and durability of concrete, including exploring innovative materials and techniques that enhance its resistance to environmental and mechanical challenges. These issues are especially evident in colder climates, where FT cycles are frequent and pose a significant challenge to the longevity of concrete use due to their effect on the structural composition. As temperature changes become more frequent due to climate change, the FT problem poses an ever-increasing societal and engineering challenge, threatening infrastructure altogether.

Freeze thaw cycles occur when water trapped inside the concrete freezes and expands during cold weather, creating structural damage and internal pressure (National Institute of Health, 2019). Once the FT cycle occurs, the concrete is prone to cracking that may increase

over time. These cracks in the concrete weaken the structure, leading to eventual failure. As the weather changes, the repeated FT cycles will undermine the concrete until it's rendered unusable. This makes it an important phenomenon to study.

A focus on sustainable building materials has also been growing, with efforts on finding ways to increase durability and lifespan with less need for repairs. Using sustainable materials contributes to environmental protection, conserves raw materials, and reduces waste. Reducing the consumption of raw materials and increasing the longevity of building materials helps minimize harmful environmental impacts. The use of nanotechnology has promising potential to lessen FT impacts, increasing the concrete's strength and longevity, and impacting sustainability.

Multi-walled carbon nanotubes (MWCNT) are a type of nanomaterial that may have an impact on FT cycles. They have very unique physical, chemical, and mechanical properties that could be used to improve concrete: they can exhibit extreme compressive, flexural, and tensile strength, as well as electrical conductivity (Filchacova, 2021). MWCNTs are arrays of cylinders made of sheets of graphene that have the same central points. Although MWCNTs have been promising, the impact that MWCNTs may have on FT resistance is unclear.

This research aims to answer how MWCNTs affect the FT resistance of concrete. Through literature reviews of experimental testing, this study seeks to fill in the gaps of how MWCNTs impact the cracking behavior and durability under FT conditions, see what improvements MWCNTs can make to concrete durability compared to regular mixtures, and identify different mixing proportions that produce the most significant benefits. By understanding the effects that MWCNTs induce in concrete, researchers are able to better understand how to create new mixtures that can avoid the damage from FT cycles in harsh environments.

Literature Review

A literature review was conducted to identify research studies that have examined the FT resistance of concrete incorporating carbon nanotubes. The focus of the literature review was on dosage, length, diameter, functionalization, and purity. Data from this literature would be organized, analyzed, and synthesized. This paper included seven papers, ranging from 2015 to 2022.

Mixture Proportions of Concrete with Multi-Walled Carbon Nanotubes

In several studies (Li *et al.* 2015; Alafogianni *et al.* 2019; Kumar *et al.* 2015; Song *et al.* 2022; Liu *et al.* 2022; Cheng *et al.* 2022) MWCNTs were mixed with cement, sand, and water. For water-to-cement ratio, three studies (Li *et al.* 2015; Kumar *et al.* 2015; Liu *et al.* 2022) had a water-to-cement(w/c) ratio less than 0.5, while the other three studies (Alafogianni *et al.* 2019; Song *et al.* 2022; Cheng *et al.* 2022) used a ratio of 0.5 or more.

Every study used paste, but every study except for Goodarzi *et al.* and Kumar *et al.* used mortar. The overall MWCNT dosage was 1% or less. Every study used a superplasticizer except for Kumar *et al.*, 2015. The range of MWCNT dosage was from 0 to 0.8 grams, with Alafogianni

et al. using the highest dosage, whereas some studies(Li *et al.*, Kumar *et al.*, Liu *et al.*, Cheng *et al.*) experimented first without MWCNTs, then gradually went up the dosage. This can be shown in Table 1.

Study	Proportions
Li <i>et al.</i> 2015	Cement - 100 g Sand - 100 g Water - 45 g MWCNT - 0-0.5 g Superplasticizer - 0.2-0.6 g W/C - 0.45
Alafogianni <i>et al.</i> 2019	Cement - 450 g Sand - 1350 g Water 225 g MWCNT - 0.2-0.8g Superplasticizer - 150% of mixture W/C - 0.5
Kumar <i>et al.</i> 2015	Cement - 99-100% of mixture Water MWCNT -0-1% by weight of cement W/C - 0.4
Song <i>et al.</i> 2022	Cement 356 kg/m ³ Sand - 691 kg/m ³ , Water - 185 kg/m ³ , MWCNT - 0.356 kg/m ³ . Superplasticizer - 1.965 kg/m ³ W/C - 0.51
Liu <i>et al.</i> 2022	Cement - 540 kg/m ³ Sand - 1080 kg/m ³ Water MWCNT - 0-0.3 kg/m ³ Superplasticizer - 2.5% of mixture W/C - 0.51
Goodarzi <i>et al.</i> 2022	Cement - 763kg/m ³ MWCNT - 0-0.3 g Water Superplasticizer - 14 lit/m ³ W/C - 0.25
Cheng <i>et al.</i> 2022	Cement - 276 kg/m ³ Sand - 707 kg/m ³ Water - 155 kg/m ³ MWCNT - 0-0.5175 kg/m ³ Superplasticizer - 10.05 kg/m ³ W/C - 0.56

Table 1: Mixture proportions in concrete with MWCNTs

Characteristics of Carbon Nanotubes

In terms of characteristics for carbon nanotubes, the length of the MWCNTs is a key factor that influences their strength and durability. Jang *et al.* shows that MWCNT lengths less than 50µm provide better dispersion in the concrete. Researchers such as Alafogianni *et al.* incorporate a MWCNT length of 10µm for their research. Jang *et al.* also shows that MWCNT lengths greater than 50µm provide better tensile strength and are more effective in bridging. In Li *et al.* the MWCNT length was between 10 to 30µm. Alafogianni *et al.* and Liu *et al.* used a MWCNT length between 5-50µm. Kumar *et al.* used a MWCNT length between 0.5 to 50µm. Song *et al.* and Cheng *et al.* used MWCNT length between 5-15µm. Goodarzi *et al.* used a MWCNT length between 20-30µm. All these can be shown in Table 2.

Study	MWCNT Length
Li <i>et al.</i> 2015	10-30 µm
Alafogianni <i>et al.</i> 2019	5-50 µm
Kumar <i>et al.</i> 2015	0.5-50 µm
Song <i>et al.</i> 2022	5-15 µm
Liu <i>et al.</i> 2022	5-50 µm
Goodarzi <i>et al.</i> 2022	20-30 µm
Cheng <i>et al.</i> 2022	5-15 µm

Table 2: MWCNT length for each study that performed the FT testing of concrete materials with MWCNTs.

The diameter plays a crucial role in the strength of the MWCNTs. Nam *et al.* states that MWCNTs with smaller diameters have enhanced the mechanical properties of the MWCNT. Li *et al.*, Song *et al.*, Cheng *et al.*, and Liu *et al.* had an MWCNT diameter between 10-20nm. Alafogianni *et al.* used a MWCNT diameter between 20-45nm. Kumar *et al.* used a MWCNT with an outer diameter between 60-100nm and an inner diameter between 5-10nm. Goodarzi *et al.* did not report a diameter. All these can be shown in Table 3.

Study	MWCNT Diameter
Li <i>et al.</i> 2015	10-20 nm
Alafogianni <i>et al.</i> 2019	20-45 nm
Kumar <i>et al.</i> 2015	60-100 nm outer diameter 5-10 nm inner diameter
Song <i>et al.</i> 2022	10-20 nm

Liu <i>et al.</i> 2022	10-20 nm
Goodarzi <i>et al.</i> 2022	N/A
Cheng <i>et al.</i> 2022	10-20 nm

Table 3: MWCNT diameter for each study that performed the FT testing of concrete materials with MWCNTs.

Specific surface area is another key parameter influencing MWCNT dispersion and reinforcement abilities. This is crucial because higher surface areas will lead to better dispersion and bonding, increasing the performance of the MWCNT in the concrete. In papers like Kumar *et al.* 2019, they had very high specific surface area between 40-300m²/g. Li *et al.* had a specific surface area of less than 120m²/g. Cheng *et al.* used a specific surface area between 100-160m²/g. Liu *et al.* had a specific surface area of 200-300m²/g. The rest of the papers did not have specific surface area recorded, as shown in Table 4.

Study	MWCNT Specific Surface Area
Li <i>et al.</i> 2015	>120m ² /g
Alafogianni <i>et al.</i> 2019	N/A
Kumar <i>et al.</i> 2015	40-300m ² /g
Song <i>et al.</i> 2022	N/A
Liu <i>et al.</i> 2022	200-300m ² /g
Goodarzi <i>et al.</i> 2022	N/A
Cheng <i>et al.</i> 2022	100-160m ² /g

Table 4: MWCNT specific surface area for each study that performed the FT testing of concrete materials with MWCNTs.

Functionalization of MWCNTs is used to improve the dispersion with other materials as well as to reinforce the durability of the concrete and allow the MWCNTs to act as “bridges” to microcracks. Right now, there isn’t much information about functionalization in the studies, so future research is needed.

Dispersion of Carbon Nanotubes

Each paper had its own way of mixing, but each of them fell into two main categories-ultrasonic dispersion or magnetic stirring. Li *et al.*, Kumar *et al.*, Song *et al.*, and Goodarzi *et al.* all use ultrasonic dispersion to mix the MWCNTs. This method utilizes high-frequency sound waves to break apart the MWCNTs, allowing the broken pieces to spread across the concrete. Other papers, like Liu *et al.*, Cheng *et al.*, and Song *et al.*, use magnetic

stirring, which uses a magnet to disperse the MWCNTs. Dispersive agents are also used to prevent clumping.

Study	MWCNT Mixing Procedure	Dispersive Agent
Li <i>et al.</i> 2015	Ultrasonic Dispersion	Superplasticizer
Alafogianni <i>et al.</i> 2019	Magnetic Stirring	Anionic surfactant; anionic polycarboxylate superplasticizer
Kumar <i>et al.</i> 2015	Ultrasonic Dispersion	Not reported
Song <i>et al.</i> 2022	Ultrasonic Dispersion/Magnetic stirring	Superplasticizer, surfactant PVP
Liu <i>et al.</i> 2022	Magnetic stirring	Surfactant PVP, Water Reducing Agent - .16-.2
Goodarzi <i>et al.</i> 2022	Ultrasonic Dispersion	Polycarboxylate
Cheng <i>et al.</i> 2022	Magnetic stirring	Polyvinylpyrrolidone K30 Surfactant(PVP)

Table 5: Mixing procedure and dispersive agents used for concrete materials with MWCNTs

Freeze-Thaw Testing of Concrete with Carbon Nanotubes

In Li *et al.* 2015, the researchers aimed to explore the mechanical properties of materials with MWCNT under FT and drying environments. First, 160 x 40 x 40mm specimens were cured in a four-day immersion in water and then spent 16 hours in a freezer followed by eight hours in water at -20°C. After that, the specimens were subjected to 30,60, and 90FT cycles. They found that there was a decrease in compressive strength reduction. Without MWCNTs, there was approximately a 19MPa loss in compressive strength, but with MWCNTs, there was approximately a 9MPa decrease in compressive strength.

In Alafogianni *et al.* 2019, the researchers explored the properties and damage of the structure after a FT process. To test this, six specimens measuring 160 mm x 40 mm x 40 mm were subjected to four hours of freezing and 300 FT cycles at a temperature of -18°C. Each cycle was four hours of -18°C and another four hours of thawing at 4°C with a humidity of 98%. They found that the compressive strength improved 10% with a maximum of 0.6% MWCNT.

In Kumar *et al.* 2015, the researchers aimed to present the effects of MWCNT in a cement paste mixture. In the study, 10 cylindrical specimens (25.4 mm x 25.4 mm x 285 mm) were first cured for 28 days and then placed in metal containers. They were subjected to 30-36 FT cycles between 0-40°F, but the specimens were limited to 5-19 cycles instead before deteriorating. They found that the relative dynamic modulus-the elasticity of the concrete, or (RDM)-decreased more rapidly than the control mix, and there was a faster rate of mass loss

compared to the control mix. This showed that the MWCNTs did not improve the cement but rather deteriorated at a faster rate.

In Song *et al.* 2022, the researchers aimed to explore how MWCNTs affect the mechanical property enhancements under heavy loads after FT cycles. The specimens were first cured for 24 days in a standard maintenance room at a maximum temperature of 22°C and 95% humidity. Then, they were soaked in water for 4 days and underwent fewer than 50 FT cycles. Each cycle lasted between 2 to 4 hours, with freezing temperature at $-18 \pm 2^{\circ}\text{C}$ and a thawing temperature of $5 \pm 2^{\circ}\text{C}$. The results show that concrete with MWCNTs showed 25.81% higher dynamic mechanical strength than regular concrete under and impact conditions. The MWCNTs also enhance the microstructure by filling it and nucleating, improving resistance to damage.

In Liu *et al.* 2022, the researchers experimented with different amounts of MWCNT and water-to-binder ratios to find the combinations that have high compressive strength. First, concrete specimens measuring 100 mm x 100 mm x 100 mm and 100 mm x 100 mm x 400 mm were demolded after 24 hours and then cured under standard conditions for 28 days with a temperature at $20 \pm 2^{\circ}\text{C}$ and 95% relative humidity. The 100 mm x 100 mm x 400 mm specimens were subsequently subjected to FT cycles ranging from $-18 \pm 2^{\circ}\text{C}$ and $5 \pm 2^{\circ}\text{C}$, with each cycle lasting approximately 10-15 minutes. A total of around 500 FT cycles were performed. The incorporation of MWCNTs resulted in a 34.6% increase in compressive strength compared to the control, with compressive strength reaching 122.96 MPa and flexural strength 14.2 MPa. This was attributed to the uniform dispersion of MWCNTs within the concrete.

In Goodarzi *et al.* 2022, the researchers studied how adding different powders such as metakaolin and fly ash along with MWCNT affects the concrete's performance. Initially, concrete specimens measuring 50cm x 10cm x 10cm were placed in a water basin with a humidity of 100% for 27 days. The specimens then underwent 300 FT cycles over 130 days with a temperature ranging from $-18 \pm 2^{\circ}\text{C}$. The experiment showed that there was a 20% loss of strength without the incorporation of MWCNT, but with MWCNT, there was only a 3% strength loss in a period of 90 days.

In Cheng *et al.* 2022, they studied how the shape and distribution of air pockets inside MWCNT-fly ash concrete influence the extent of damage to the concrete when it experiences FT. Initially, after completing standard curing, the concrete samples were immersed in a sodium sulfate solution for 28 days at a temperature range of $20 \pm 2^{\circ}\text{C}$ at 95% relative humidity. Then, the concrete samples were soaked for 4 days and were transferred to a FT chamber set at -20°C , with FT testing from 0 to 200 cycles in increments of 25. However, since the MWCNTs didn't space themselves well, they did more harm than good, so there will be additional future tests.

Discussion Analysis

The usage of the MWCNTs into concrete has overall shown very good results in improving the concrete's FT resistance. Most projects show positive results of how the usage of MWCNTs had led to significant benefits and an improvement to the concrete's FT resistance. In various studies, positive effects were seen, as Li *et al.* 2015 experienced lower losses in

compressive strength and also found that the MWCNTs have a good effect of bridging and reducing microcracks. Their experiment had a dosage of 0-0.5g with a length between 10 and 30nm, diameter between 10 and 20nm, and specific surface area of less than 120m²/g. These may provide scientists a general sense of the desired characteristics and dosage of the MWCNT for better results. Similarly, Alafogianni *et al.*, Song *et al.*, Goodarzi *et al.*, and Liu *et al.*, all showed desired results of the FT, including enhanced compressive strength and reduction of cracking, thereby showing the potential of MWCNTs to be used in the enhancement of concrete.

An important key factor that led to these benefits was the dispersion of the MWCNTs. The more evenly dispersed the MWCNTs are, the more equally the benefits of the MWCNT apply to each part of the concrete. This uniformity reduces the chances of weak areas and contributes to the overall improvement of the concrete structure. Dispersion techniques such as magnetic stirring and ultrasonic dispersion helped optimize the distribution of MWCNTs, making the concrete become more robust. Magnetic stirring occurs when a magnetic stir bar is dropped in a solution, causing the liquid to mix inside while also dispersing the MWCNTs. Ultrasonic dispersion occurs when high-frequency soundwaves break clumps of chemicals apart, allowing more evenly spread dispersions.

However, some results show very different results, as in Kumar *et al.* 2015 and Cheng *et al.* 2022, where both effects are negative due to the lack of dispersion of the MWCNT, as the experiment in Kumar *et al.* deteriorated faster with the MWCNT than the control group that doesn't include the MWCNT, which indicates that the MWCNT may have negatively impacted the concrete. Kumar *et al.* did not use any surfactants when dispersing the MWCNT which may have played a role in their results. In Cheng *et al.*, researchers found that there were various inconsistencies within the concrete, damaging the overall compressive strength. These contrasting results show the importance of controlling the dispersion and mixing procedures to maximize the effectiveness of MWCNTs.

Additionally, lower diameters and higher length, specifically diameters less than 25nm and lengths around 30μm, show the overall best results. The ideal water to cement ratios around 0.5 are also the ideal dosage, whereas the MWCNT to concrete ratio around 0.005 maximizes the potential of the concrete.

Unfortunately, for functionalization, only Li *et al.* used it, whereas the other studies didn't. Future research is needed to discover the effects functionalization has on the FT resistance.

Conclusion

In this study, the impacts of MWCNTs in FT resistance of concrete was studied. A total of seven papers from 2015-2022 were analyzed. The MWCNTs showed significant potential on impacting the usage of concrete through the improvements of FT resistance and mechanical properties of compressive strength and durability. The research also suggests that MWCNTs with smaller diameters and lengths were found to be the most effective, along with water to concrete ratio of 0.5 and a MWCNT to cement ratio of 0.005.

However, there were challenges to achieve the desired results, with several factors playing a role in the results. While most papers showed benefits, the dispersion of the MWCNTs became a problem to some studies that negatively impacted the MWCNT-enhanced concrete, resulting in decreases in compressive strength and durability. Future work should be dedicated to optimizing the carbon nanotube characteristics and mixing procedure for each type of concrete material so that the desired FT resistance can be achieved.

The results of this study show the growing potential MWCNTs may have in the construction field and may play a significant role in the future, creating more durable and sustainable infrastructure. Improving dispersion techniques may help the performance of the MWCNTs in different environmental conditions.

Works Cited

- Li, Wei-Wen, et al. "Investigation on the Mechanical Properties of a Cement-Based Material Containing Carbon Nanotube under Drying and Freeze-Thaw Conditions." MDPI, publisher, 14 Dec. 2015, www.mdpi.com/1996-1944/8/12/5491.
- Alafogianni, Panagiota, et al. "Structural Properties and Damage Detection Capability of Carbon Nanotube Modified Mortars after Freeze-Thaw." MDPI, publisher, 29 May 2019, www.mdpi.com/1996-1944/12/11/1747.
- Kumar, Sanjeev, et al. "Effect of Multiwalled Carbon Nanotube in Cement Composite on Mechanical Strength and Freeze-Thaw Susceptibility | Advances in Civil Engineering Materials | ASTM International." Effect of Multiwalled Carbon Nanotube in Cement Composite on Mechanical Strength and Freeze-Thaw Susceptibility, dl.astm.org/acem/article-abstract/4/1/257/257/Effect-of-Multiwalled-Carbon-Nanotube-in-Cement?redirectedFrom=fulltext. Accessed 28 Dec. 2025.
- Song, Shushuang, et al. "Study the Effect of Carbon Nanotube on Mechanical ..." Study on Dynamic Mechanical Properties of Carbon Nanotubes Reinforced Concrete Subjected to Freeze–Thaw Cycles, [ijaem.net/issue_dcp/Study the Effect of Carbon Nanotube on Mechanical Properties of Concrete.pdf](http://ijaem.net/issue_dcp/Study%20the%20Effect%20of%20Carbon%20Nanotube%20on%20Mechanical%20Properties%20of%20Concrete.pdf). Accessed 28 Dec. 2025.
- Liu, Guifeng, et al. "Experimental Study on the Salt Freezing Durability of Multi-Walled Carbon Nanotube Ultra-High-Performance Concrete." MDPI, publisher, 28 Apr. 2022, www.mdpi.com/1996-1944/15/9/3188.
- Iman Goodarzi, S. Mohammad Mirhosseini & Ehsanollah Zeighami 2022 Experimental study of the freeze and thaw effects on the mechanical behaviour of self-compacting high-performance fibre-reinforced concrete containing active powders and/or Carbon NanoTubes, *Australian Journal of Structural Engineering*, 23:1, 37-50, DOI: 10.1080/13287982.2021.1998994
- Cheng, xu, et al. Grey Entropy Analysis of Strength and Void Structure of Carbon Nanotubes Concrete under the Coupling of Sulfate Attack and Freeze-Thaw Cycles, 2022, www.sciencedirect.com/science/article/abs/pii/S095006182203118x.

Holding Up a Generation: Reimagining How Schools, Parents, and Peers Support Teen Mental Health

By Jiya Aghi, Kaeya Aghi, Sonia Jain

Keywords

Mental health, teen mental health, ADHD, PTSD, trauma, school, education system, parents, peer support

Abstract

Teen mental health has emerged as one of the defining public health challenges of our generation. Rates of anxiety, depression, ADHD, PTSD, loneliness, and academic stress have risen sharply, with girls, neurodivergent youth, and those navigating invisible disabilities disproportionately affected. This review synthesizes interdisciplinary research from psychology, education, neuroscience, and public health with real-world observations of youth with complex mental health needs and their families navigating underprepared, rigid systems with devastating impacts on their functioning, learning and health. Together, these insights illuminate how school systems, family environments, and peer dynamics can both harm and protect adolescent well-being.

Evidence consistently shows that trauma-informed school practices, supportive parental relationships, and healthy peer connections are among the strongest protective factors for teen mental health. Yet structural gaps in schools - such as understaffed counseling departments, inadequate teacher and school staff training, rigid academic policies, variable teacher expectations, and dismissive communication between schools and families - create major preventative barriers that undermine student resilience and success. This review outlines the unique struggles of adolescents with ADHD and trauma, and the importance of belonging, empathy, relational safety, and the systemic changes urgently needed to build supportive environments across schools, families, and teen communities throughout the US. Recommendations emphasize practical, evidence-based strategies to cultivate greater understanding, capacity and cultures of compassion, connection and collaborative holistic mental health supports and services.

Introduction

Teen mental health has quietly become one of the defining public health epidemics of our generation. Many youth ages 10-24 struggle silently with invisible mental health conditions like anxiety, depression, trauma, Attention-Deficit Hyperactivity Disorder (ADHD), loneliness, and academic pressures. Nationally, teen mental-health challenges are broad, rising among adolescents of all races and socioeconomic backgrounds, in both more privileged and underprivileged communities (Xiang, 2024). Rates of anxiety, depression, ADHD, trauma, and stress have risen sharply in the past few years, especially among girls, neurodivergent youth, and those navigating invisible disabilities disproportionately affected. Whether formally diagnosed, borderline, or undiagnosed, many teens suffer unfairly internally and in their social lives.

According to the Centers for Disease Control (CDC, 2023), one in three teen girls reports persistent sadness or hopelessness. These numbers rise even more for those who are already managing attention difficulties, trauma histories, or chronic stress – often misunderstood by teachers, school staff, and peers.

Behind every statistic is a young person trying to make it through the school day while balancing overwhelming parental or societal (e.g. academic) expectations, shifting dynamic friendships, constant digital comparison and distractions and uncertainty about the future. Our generation is the first to grow up in a fully digital world with internet and devices, social media and structured activities, along with less play, and more isolation, significantly affecting our attention span, mental health, and socioemotional development (Twenge, 2017; APA, 2023; WHO 2022). For many teens with compromised mental health, school is the main place where their struggles collide with systems that are not built for the complexity of their needs, and they must daily navigate the power dynamics of underprepared adults in positions of power, and unempathetic peers while managing their deeply personal mental health conditions.

This paper highlights interdisciplinary research from psychology, education, neuroscience, and public health, and blends it with real-world observations from a high schooler witnessing the growing mental health crises and the impacts of limitations in our school system, friends, and community. The goal is to raise greater awareness of this issue, which has a detrimental effect on students' trajectories, and to bring about change in schools, peers, and communities globally. It is a pivotal time in our history that schools, families, and peer communities must transform traditional systems and create inclusive, equitable, supportive school climates and cultures where all teenagers, especially those with anxiety, depression, ADHD, trauma, disabilities and other comorbidities, feel supported and thrive.

The State of Teen Mental Health

Over the past decade, it has become evident that teen mental health is worsening dramatically. Anxiety, depression, ADHD, PTSD, and loneliness or disconnection among teens continue to rise (APA, 2022), and we are still not talking about it. Around 40% of teens experience significant anxiety symptoms (Twenge, 2023), and one in five meets criteria for a mental health disorder (NIMH, 2023). It is most likely underreported and underdiagnosed due to stigma still associated with it, high cost and limited use of mental health services, cultural norms among diverse racial/ethnic and immigrants growing the US, and limited capacity of school-based counselors and other socio-ecological shifts happening affecting teen life. Suicide remains the second leading cause of death for adolescents (NIMH, 2022). It is a pandemic with detrimental effects on today's youth, compromising the human development of millions of teens, including their physical, emotional, social, and cognitive development.

For girls and neurodivergent youth, these pressures often collide with conditions that go unnoticed. ADHD in girls, for example, is frequently missed or mistaken for apathy or a lack of motivation (Quinn & Madhoo, 2014; Almekhlafi & Jain, 2024). ADHD in girls is often missed or misinterpreted because many present primarily with inattentive or internalized symptoms

(e.g., difficulty concentrating, disorganization, forgetfulness) rather than overt hyperactivity or impulsivity, which is more common in boys, leading to misunderstandings, underdiagnosis, or misdiagnosis as apathy, laziness, or lack of motivation. PTSD, often thought of as a response to one big event, can also stem from ongoing stressors such as ongoing bullying prevalent during transitional middle-school years, unstable home environments, lack of community, racist incidents, or sudden losses (NCTSN, 2020). Studies show trauma exposure is especially high among young women (Dimitry, 2020).

When ADHD and trauma interact, symptoms intensify (Njardvik et al, 2025; Magdi, 2025). This is an understudied population, as many teens with trauma do not seek or receive services, and ADHD goes undetected. However, we can estimate that roughly 1 in 3 teens or 28-36% of teens with ADHD also show PTSD symptoms or diagnosis, though this may be lower in the general population (Magdi, 2025). A large meta-analysis of 121 studies found that nearly 39,894 children and adolescents with ADHD had comorbidities such as behavioral disorders, anxiety, OCD, etc. These students, especially girls, with invisible symptoms may struggle with focus, memory, organization, emotional regulation, and social relationships. They may also develop other mental health conditions not well understood like PICA which may further stigmatize or isolate the teenager, with judgements from peers and school staff. These teens may seem withdrawn, overwhelmed, or inconsistent - behaviors often misunderstood by adults around them. This should trigger concern and care, not judgement and victim-blaming.

In parallel, an unprecedented rise in disconnection, reduced empathy and weakened community bonds has emerged, partly as a result of growing digital immersion during and post COVID-19 pandemic. Research shows that adolescents today spend significantly more time online and less time in face-to-face interaction, contributing to declines in empathy, social skills, and emotional attunement (Konrath et al., 2011; Twenge, 2023). Particularly during middle and high school years when social connections and fitting in with friends is paramount, peer relationships undergo significant shifts, and many youth experience declines in perceived empathy and emotional support from friends, feeling more isolated, a pattern more pronounced among adolescents with mental health conditions (Rueger, Malecki and Demaray 2010; Orben and Tomova and Blakemore, 2020).

In addition, students with compounded mental health conditions must continue to unfairly navigate school systems and staffing that still largely operate within traditional, “old-school” structures with limited understanding and capacity to recognize or respond to these misunderstood complex mental health symptoms that visibly appear as attention difficulties, sensory overload, digital addiction, apathy, social withdrawal, and trauma-related stress. The combination of rapidly shifting adolescent social worlds and rigid systems that have not evolved at the same pace has created a quiet *parallel pandemic*, a growing crisis of unmet teen mental health needs and emotional disconnection with detrimental impacts on teen lives. This phenomenon warrants far greater attention, investment, and intervention across education, public health, and community settings before its impacts become even more deeply entrenched (APA, 2022; Murray et al., 2020). We have reached a critical moment in time, where we urgently need

to rethink and reform how we support teens, not through punishment or pressure, but by creating welcoming environments where teens feel safe, understood, and supported.

Schools: A Source of Harm or Healing?

Schools play a critical role in the whole adolescent development and success. Many students with ADHD or trauma histories are falling through the cracks, as many school systems are unprepared and re-traumatizing and harming the students, compromising their learning and development, instead of healing or helping. Many such students are unintentionally labeled and treated as “lazy,” “defiant,” “unmotivated,” or “apathetic” by adults and peers at schools. Strict deadlines, zero-flexibility academic policies, and punitive discipline further isolate those already struggling. Even front-office interactions shape whether a student feels safe or dismissed. No doubt schools across the country are stretched thin. Many school systems are under- or inadequately prepared, even in privileged areas, for the rising complex mental health needs of students today. This is due to insufficient acknowledgement at national, state or local levels, insufficient staffing and resources, high staff or counselor turnover, limited staff and teacher training and capacity, and rigid academic policies.

Common school system gaps include:

First, Insufficient mental health staffing and high turnover, partly due to low pay. The recommended counselor-to-student ratio is 1:250; the U.S. average is 1:430 (ASCA, 2022), and the Psychologist-to-Student ratio is 1:1,100 (NASP, 2023). School counselors are usually the first point of contact for students with social-emotional concerns and their families. Counselors are supposed to help connect students to support and manage crises. At the same time, school psychologists provide assessments and expertise in trauma-ADHD. However, both are limited by large caseloads, high turnover, and a limited understanding of complex symptoms, leading to inappropriate follow-ups with devastating impacts on the student.

Second, school staff and teachers, though well-intentioned, have minimal training in ADHD, trauma, and adolescent mental health; thus, a limited understanding of symptoms, which has implications for student’s attendance, school connection, engagement and performance. In fact, this leads to judgment, stigmatization, and re-traumatization of the student and family. Teachers often notice behaviors if something is wrong or if the student is not behaving in the classroom according to their rules or expectations. And trauma-informed teaching—recognizing signs of distress, avoiding punishment, creating predictable routines, offering flexible deadlines, using UDL, and speaking in neutral language—can make school feel safer and more manageable for struggling students (Dorado et al., 2016). Teachers often receive under 10 hours of trauma training (Atkins et al., 2020). Misinterpretation of symptoms is common and leads to stigmatization.

Teacher training and practices in trauma-informed care are minimal, with detrimental impacts on the youth, leading to misinterpretation of mental health symptoms and limited support. Rather, it leads to judgment, isolation, and fear-based responses to students with ADHD/PTSD and the like. Rigid tests, assignment deadlines, and inflexible grading structures

overwhelm students with ADHD or trauma histories; yet these are dependent on teachers' personalities and decisions, who may have limited empathy, compassion, and understanding of those students' behaviors. Front office staff, including counselors, vice principals, attendance and principal, play a critical role in setting and enforcing supportive policies and procedures, school culture and climate, and educating peers and teachers on an ongoing basis, often lack understanding and trauma-informed training and practices as well, even with good intentions. This leads to insensitive communication, a lack of support, and more stigma and unfair treatment of the students with mental health issues.

Third, there are rigid academic structures and rules, school-wide and within classes, that teachers do not generally bend. Strict deadlines and heavy workloads penalize students with executive functioning difficulties (Graham et al., 2018).

Moreover, fourth, administrative and front-office barriers are paramount. Office staff shape school climate but receive no training in sensitive communication (Bryk & Schneider, 2002). Administrators, especially vice principals, strongly influence school culture; when they lack understanding of trauma-informed practices, their decisions can unintentionally harm students. Leadership rooted in kindness, flexibility, restorative practices, and respectful communication with families is essential (Chafouleas et al., 2016). Even front office and attendance staff shape students' daily experiences, and without training in empathy, confidentiality, and mental-health awareness, interactions can feel judgmental or unfair, especially for students managing chronic conditions. Indeed, there needs to be a lot more research to examine the role of front office staff and trauma-informed training, attitudes, practices, and sensitivity to students with ADHD/PTSD and complex disorders. We know when schools adopt trauma-informed and neurodiversity-affirming practices, i.e., predictable routines, flexible grading, restorative approaches, Universal Design for Learning, and collaborative communication—students' mental health outcomes significantly improve (Dorado et al., 2016; SAMHSA, 2014). True school transformation hinges on something simple but powerful: Every adult understands the impact of trauma, attention challenges, and chronic stress—and treating students with empathy, patience, and respect.

How can students with ADHD/PTSD and other complex invisible mental health issues feel safer, more connected, and supported by school staff, teachers, and peers in traditional school settings? Schools in the US could benefit from ensuring implementation of trauma-informed practices, based on SAMHSA's six principles (2014). These include safety, trustworthiness, peer support, collaboration, empowerment, and cultural sensitivity. How to develop a more flexible academic curriculum and structure (e.g., hybrid models, flexible deadlines, more tutoring)? These include flexible deadlines, grading and assessment, workload modification, trauma-informed communication, predictable structures, and strengthened home-school collaboration. A coordinated school team, but most importantly, each school staff and teacher's sensitivity, empathy, understanding, and treatment of the students with social-emotional concerns is paramount. Finally, evidence-based interventions such as MTSS, SEL programs, mental health screenings, and strong school-family partnerships should be

annually evaluated and improvements made by the schools to strengthen students with mental health needs, resilience, and growth within traditional school settings. There is a lot we can do and must be done if we want to ensure the success of a generation that is breaking down.

Parents as Frontline Advocates and Support

Parents are often the first to recognize changes in a teenager's emotional or behavioral patterns at home. But families managing ADHD, PTSD, anxiety, or depression, especially among their teens, often encounter significant barriers and challenges repeatedly at every level within schools, districts, or mental health systems. Many feel judged, dismissed, or inadequately supported, caught in a bureaucratic run-around, navigating rigid policies, fragmented procedures, that delay or prevent timely, compassionate support (Brye & Schneider, 2002; Murray et al, 2020). Parents often understand their teenager's condition in ways no one else can notice subtle shifts, triggers, and patterns - information essential for effective support, if taken seriously. However, despite approaching schools with trust and respect, their voices are often undermined and undervalued, despite repeatedly advocating for adequate ongoing support and collaboration, delaying effective interventions (Bryk & Schneider, 2002; Murray et al, 2020).

A student may be struggling in PE due to chronic fatigue, PTSD triggers, or medication side effects. A parent may request flexibility and be told, "We treat all students the same." This lack of nuance can worsen symptoms, attendance, and motivation. Schools need to listen more closely to parents and make informed, nuanced decisions throughout the day to ensure students are safe and appropriately accommodated. This includes class-specific support—such as a PE teacher allowing limited participation for a student with chronic fatigue syndrome or PTSD (Murray et al., 2020). Parents also need support systems or groups in place, with other parents facing similar issues and navigating such complex, ineffective systems. Parents need to be treated as collaborative partners, not as obstacles or overprotective caretakers. Trauma-informed school-family relationships—built on empathy, transparency, and shared problem-solving—have been shown to significantly improve student outcomes (Murray et al., 2020).

Parental warmth, consistency, and emotional availability are among the strongest predictors of resilience in teens (Rogers, 2021). Families themselves also need support. Navigating inconsistent mental health systems, long diagnosis waitlists, or mismatched therapists can be overwhelming. Parent support groups, skill-building workshops, and community networks can provide crucial relief and guidance.

Therapy and Treatment: What Teens Really Need

Mental health is a prerequisite for cognitive functioning, learning, academic performance, proper behavior in classrooms, and social engagement with 'normal' peers. Therapeutic milieu is essential within a school environment. What teens really need is not only proper assessment, but also individualized mental health therapy and treatment, including for borderline youth. The online mental health therapy, mindfulness practices, and rise of physical exercise and 'health' overall are great options for teens today. However, the stigma to access and using mental health

therapy is still prevalent. The landscape is shifting slowly. While therapy is essential, the treatment landscape is fragmented. Research shows that up to 50% of teens stop therapy because of poor fit to the dynamic and complex needs, high costs, or lack of progress (Warden et al., 2020).

Additionally, school psychologists or community-based therapists often lack specialized training in ADHD/trauma or comorbid symptoms. And ongoing coordination with parents, essential for meaningful interventions at school, is frequently limited. Therapy must be trauma-informed, culturally responsive, developmentally appropriate, and collaborative, yet many providers rely on general approaches rather than disorder-specific care. Greater training and preparation of psychologists who understand and can deal with the nuances of comorbid symptoms within the context of teen lives and culture is absolutely critical. Because teens' symptoms fluctuate daily, therapists cannot rely solely on the youth's perspective; with appropriate attention to confidentiality, regular communication with parents and schools is essential for accurate assessment and aligned treatment.

Therapists may see a teen weekly but have limited understanding of how symptoms show up in school. And because teens' moods and attention vary day to day, relying solely on their self-report can lead to incomplete or inaccurate treatment plans. The most effective interventions such as CBT, DBT, Trauma-Focused CBT, organizational skills training, mindfulness, and movement-based practices work best when they are connected to and built into the teen's daily environment and supported by caregivers and educators (Zenner et al., 2014; Almekhlafi et al., 2026). Though there is a constant art of push and pull and effectively engaging and involving youth with ADHD-PTSD. Others have noted and we have observed that teens with ADHD demonstrate improved functioning in high-activity, high-stimulation environments that provide greater stimulation, structure, or time pressure (Hallowell & Ratey, 2021)

Mindfulness-based interventions are also increasingly accessible and have demonstrated benefits for improving attention, awareness, and emotional regulation, particularly among adolescents (Almekhlafi et al., 2026). Regular physical activity, whether through structured exercise or movement incorporated at home or school, has consistently been shown to improve teen mental health and reduce stress. Creative outlets such as baking, art, and music further support concentration, emotional expression, and overall well-being, offering meaningful pathways for teens with social-emotional concerns to regulate and thrive.

Peers, Empathy, and the Importance of Belonging

Peer relationships can make or break a teenager's mental health. Supportive friendships reduce the risk of depression, anxiety, and loneliness (Bukowski et al., 2018). For teens with ADHD or PTSD, the stakes are even higher. These teens are often more sensitive to rejection and rely heavily on close, trusting relationships. Unfortunately, many constantly face exclusion or isolation, bullying, misunderstanding, and social exclusion or subtle relational aggression at many levels. Research shows that social exclusion activates the same part of the brain as physical pain (Eisenberger, 2012), and many teens with mental health challenges endure this pain daily.

At a time when they most need support from friends, many such teens instead face judgment, misunderstanding, and a lack of empathy, further aggravating emotional distress and undermining mental health (Bukowski et al., 2018; Eisenberger, 2012). For teens with ADHD or trauma histories, fostering positive peer interactions is especially crucial—caring friends who check in, understand their struggles, and offer consistent, genuine support. Thus, friends’ parents, teachers, and adults can play a major role in facilitating a supportive environment by peers in the classroom or extracurricular after-school activities.

Belonging, that is feeling known, valued, and connected, is one of the strongest protective factors for adolescent well-being (Allen et al., 2018; Resnick et al., 2010). Even one caring adult can dramatically change a teens’ life trajectory and show reduced signs of depression and anxiety. This has been shown again and again in resilience research, the power of protective factors (Jain et al, 2011). Predictable, supportive relationships can significantly improve emotional regulation, and thereby performance and development for students with ADHD/PTSD and comorbidity. If schools can actively and intentionally cultivate relationships rooted in kindness and connection, student mental health improves—even when academic pressures remain.

The Need for Human Kindness and Connection

There seems to be declining empathy among teenagers and adults alike at school and in communities. Konrath et al. (2011) found that self-reported empathy decreased by 40% over three decades, influenced by digital life, academic pressure, and reduced face-to-face interaction. Teens today are overstimulated and overscheduled. Social media intensifies comparison, which can lead to jealousy, self-criticism, and increased anxiety (Vogel et al., 2014). When emotional bandwidth shrinks, empathy is often the first thing to go. This is an essential part of whole human development with massive future implications on our society and the world. We need to create intentional opportunities for our youth to connect face-to-face, develop social skills, and for shared vulnerability without stigma, judgment, or bullying. Again, schools and parents can play a leading role in this shift by creating environments where compassion is modeled and reinforced—not just expected.

Discussion

Adolescent mental health is shaped by the systems around them. When school policies, parental support, and peer culture align with the developmental needs of teens, they flourish. When those systems are rigid, under-trained, or unresponsive, even the strongest young people will struggle. The complexities of modern adolescence require coordinated responses from schools, families, and communities. ADHD, PTSD, anxiety, and depression intersect with school structures, teacher expectations, and peer culture in ways that either support or hinder students’ mental health. Trauma-informed and neurodiversity-aware practices must not be optional enhancements anymore, but essential components of academic environments, adequately resourced and implemented in every school in the country.

Schools must shift from punitive approaches to supportive ones. A growing body of evidence shows that students thrive when surrounded by predictable, flexible routines, relational safety, and adults who understand the science of trauma and attention. Specifically, stronger teen mental health necessarily requires: 1) trauma-informed flexible school environments: Predictability, safety, UDL, and adjusted deadlines help students with attention or trauma needs. 2) coordinated school teams, that is, teachers, parents, counselors, psychologists, and administrators, must collaborate regularly to form ecosystems and a continuum of care and learning. 3) Empowered families and youth, where parents and teens feel heard, respected, and included in all levels of decision-making regarding their students. Parents need communication, resources, and meaningful partnership. 4) Individualized therapy connected to daily life, for example, CBT, DBT, TF-CBT, mindfulness, and movement practices, work best when therapists, schools, and families coordinate. 5) Cultures of empathy and connection - Students need at least one adult, and ideally many, who absolutely believe in them. And finally, 6) Policies and staff /leadership practices in trauma-informed mental health staffing and capacity. Communities must advocate for policies that allocate funding and staff to mental health. Without adequate counselors, psychologists, or trauma-trained educators and school practices, all other progress is limited.

In the end, teen mental health is not about “fixing” young people or of individual weakness, but a reflection of systems that are misaligned with adolescent needs. It is about building systems that understand them and collectively carefully address their needs and achieve whole student development. Schools must shift from punitive to supportive frameworks. Parents need better tools and stronger partnerships. Peers must learn empathy and relational responsibility.

A reimagined school model - one that is trauma-informed, flexible in assessment and comprehensive supports, empathetic communication and culture, and nuanced daily interventions - can transform outcomes for millions of students. Adolescents deserve environments where they are treated as whole people, not data points. They need teachers trained in mental health literacy, administrators prioritizing well-being, and peers guided toward compassion. They deserve environments that recognize their complexity and honor their potential. When we ensure school systems, peers and communities are rooted in safety, belonging, flexibility, compassion, and connection, teens worldwide will not just cope - they will thrive.

Works Cited

- Allen, K. A., Kern, M. L., Vella-Brodrick, D., Hattie, J., & Waters, L. (2018). What schools need to know about fostering school belonging: A meta-analysis. *Educational Psychology Review*, 30(1), 1–34.
- Almekhlafi, K., and Jain, S. (2024) Unveiling Gender Disparities in ADHD: A Literature Review on Factors and Impacts of Late Diagnosis in Females (2010-2023). *Journal of Women's Mental Health* - 1(1):9-21.
- Almekhlafi, K., Macdonald, H., Malik, K., and Jain, S. (2026, Under Review). Effects of Mindfulness-Based Interventions on the Mental Health of Adolescent Girls and Young Women: Scoping Literature Review 2015-2025. *Journal of Research on Adolescence*.
- American Psychological Association. (2022). *Stress in America™ 2022: Concerned for the future, beset by inflation*.
- American School Counselor Association. (2022). Student-to-school-counselor ratio 2021–2022.
- Atkins, M. S., Sheridan, M., & Li, K. (2020). Mental health and trauma-informed training gaps in schools. *Journal of School Health*, 90(8), 654–662.
- Barkley, R. A. (2018). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (4th ed.). Guilford Press.
- Bryk, A. S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. Russell Sage Foundation.
- Bukowski, W. M., Laursen, B., & Rubin, K. H. (2018). Peer relations and psychopathology. *Annual Review of Psychology*, 69, 133–159.
- Centers for Disease Control and Prevention. (2023). *Youth risk behavior surveillance—United States, 2021*.
- Chafouleas, S. M., et al. (2016). Trauma-informed schools: A framework for interventions. *School Psychology Quarterly*, 31(3), 327–340.
- Dimitry, L. (2020). Global prevalence of mental health disorders in youth. *Journal of Adolescent Health*, 67(2), 145–152.
- Dorado, J. S., et al. (2016). Healthy environments and response to trauma in schools. *School Mental Health*, 8(1), 163–176.
- Eisenberger, N. I. (2012). The pain of social disconnection: Examining the shared neural basis of physical and social pain. *Nature Reviews Neuroscience*, 13(6), 421–434.
- Gathercole, S. E., & Alloway, T. P. (2008). *Working memory and learning: A practical guide for teachers*. Sage.
- Graham, S., et al. (2018). Executive function challenges in academic settings. *Learning and Individual Differences*, 67, 189–200.
- Jain, S., Buka, S. L., Subramanian, S. V., & Molnar, B. E. (2012). Protective factors for youth exposed to violence: Role of developmental assets in building emotional resilience. *Youth Violence and Juvenile Justice*, 10(1), 107–129. <https://doi.org/10.1177/1541204011424735>
- Konrath, S. H., O'Brien, E. H., & Hsing, C. (2011). Changes in dispositional empathy in American college students over time. *Personality and Social Psychology Review*, 15(2), 180–198.

- Masten, A. S. (2014). *Ordinary magic: Resilience in development*. Guilford Press.
- Murray, C., Finigan-Carr, N., Jones, K., Copeland-Linder, N., Haynie, D. L., & Cheng, T. L. (2020). Barriers and facilitators to school–family partnerships for youth mental health. *Journal of School Health*, 90(10), 748–757. <https://doi.org/10.1111/josh.12930>
- National Association of School Psychologists. (2023). Shortages in school psychology: Challenges and consequences.
- National Child Traumatic Stress Network. (2020). Understanding child traumatic stress.
- National Institute of Mental Health. (2022). Mental health statistics for adolescents.
- Quinn, P. O., & Madhoo, M. (2014). A review of attention-deficit/hyperactivity disorder in women and girls. *Current Psychiatry Reports*, 16(7), 1–8.
- Resnick, M. D., et al. (2010). Protective factors, resiliency, and healthy youth development. *Journal of Adolescent Health*, 48(3), 1–8.
- Substance Abuse and Mental Health Services Administration. (2014). SAMHSA’s concept of trauma and guidance for a trauma-informed approach.
- Substance Abuse and Mental Health Services Administration. (2019). Behavioral health of youth in the United States.
- Twenge, J. M. (2023). *Generations: The real differences between Gen Z, millennials, Gen X, boomers, and silents*. Atria Books.
- Vogel, E. A., Rose, J. P., Roberts, L. R., & Eckles, K. (2014). Social comparison, social media, and self-esteem. *Psychology of Popular Media Culture*, 3(4), 206–222.
- Warden, D., et al. (2020). Factors associated with psychotherapy drop out in adolescents. *Journal of Child and Adolescent Psychopharmacology*, 30(3), 161–169.
- Zenner, C., Herrnleben-Kurz, S., & Walach, H. (2014). Mindfulness-based interventions in schools: A systematic review and meta-analysis. *Frontiers in Psychology*, 5, 603.
- Danielson, M. L., et al. (2024). Prevalence of ADHD among U.S. children and adolescents. *Journal of Child Psychology and Psychiatry*.
- Gelner, H., et al. (2023). Effects of PTSD and ADHD symptom interaction on psychosis-like experiences. *Frontiers in Psychiatry*, 14, 1232606. <https://doi.org/10.3389/fpsy.2023.1232606>
- Magdi, H. M., et al. (2025). Attention-deficit/hyperactivity disorder and post-traumatic stress disorder: A systematic review of comorbidity across the lifespan. *Systematic Reviews*, 14, 87. <https://doi.org/10.1186/s13643-025-02774-7>
- Njardvik, U., Wergeland, G. J., Riise, E. N., Hannesdóttir, D. K., & Öst, L. G. (2025). Psychiatric comorbidity in children and adolescents with ADHD: A systematic review and meta-analysis. *Clinical Psychology Review*, 118, 102571. <https://doi.org/10.1016/j.cpr.2025.102571>
- Xiang, A. H., Martinez, M. P., Chow, T., et al. (2024). Depression and anxiety among U.S. children and young adults. *JAMA Network Open*, 7(10), e2436906.

Existing Methods for Managing Fear, Anxiety, and Stress in Shelter Animals

By Nancy Sun

Abstract

The prevalence of Fear, Anxiety, and Stress (FAS) is a major issue in animal shelters. The shelter environment is an unfamiliar and unpredictable space, and staying at the shelter often heightens an animal's stress. FAS is detrimental to both physical and mental welfare of shelter animals, worsening behavior problems or even leading to euthanasia. Therefore, it is essential to notice FAS in animals, respond to it, and take all measures to reduce it. By reviewing the scientific literature and guidelines from animal welfare organizations to compile resources on FAS, this study aims to find what measures should be taken to reduce fear, anxiety, and stress in animal shelters. Section 1 is about recognition and response to FAS, and it discusses behavioral signs of fear, behavioral cues and biochemical cues of FAS, behavioral scoring scales that can aid in FAS assessment, distinguishing and managing pain, and methods for behavior modification. Section 2 focuses on environmental factors, including design considerations for the enclosure and environment, general housing and enrichment guidelines, benefits of co-housing, and the effects of appeasing pheromones in the shelter setting.

Introduction

Fear, Anxiety, and Stress (FAS) is commonly found in shelter animals, as the shelter environment is often a place of general discomfort. Overstimulation from strong smells or loud sounds, long periods of confinement, limited social interaction with people, lack of choices in their environment, and unpredictability of a shelter setting are all factors that worsen the animal's stay at the shelter, according to the The Association of Shelter Veterinarians or ASV. This causes medical and behavioral problems to emerge or worsen due to the animal's stressed state, and the animal is unable to cope due to the limitations of their confinement. Animals in a stressed state are more prone to illness, as the immune system would be compromised. Stress may also induce or worsen behavior issues, which are a common reason for animal relinquishment (Levy et al. "Module 8: The Case of the Barking Dog"), which would result in more shelter backlog and euthanizations. It would also contribute to the working conditions and overwork of shelter staff, who experience high levels of stress and burnout (Wolf et al. 6-7).

The purpose of this study was to identify ways to recognize and respond to FAS behaviors, and outline environmental adjustments for reducing FAS. This was done by reviewing existing literature on reducing FAS in shelters and for animals in general.

Section 1 of this article discusses methods to recognize FAS and reduce it using biochemical cues or behavioral assessment tools, pain scales to distinguish stress from pain, and examples of behavior management protocols. Reducing FAS is essential to prevent animal reactivity and improve welfare in the shelter. To reverse or correct behavioral issues, it is better to intervene sooner than later to prevent escalation of reactivity. Many animals in the shelter environment may be experiencing physical pain, which is difficult to distinguish from stress as

they are often co-occurring. It is necessary to determine when animals are experiencing pain, as there could be underlying medical needs which affect their quality of life.

Section 2 discusses the environment and routine of the shelter, which also influence the animal's comfort during their stay and their FAS level. Factors of housing such as enclosure size, ambient noise, sharp smells, temperature, ventilation, and lighting should be considered for minimizing discomfort. The section will also discuss the importance of kennel setup, regular enrichment and interaction with people, and co-housing with compatible animals. Use of synthetic appeasement pheromones products such as Adaptil and Feliway will be touched on as well. These products replicate the appeasing pheromones that mother dogs or cats release after giving birth to strengthen the bond with their young and are meant to have a calming effect for conspecifics, but their effectiveness is debated (PetMD Editorial and Coates; Chadwin et al. 418).

Section 1: Recognition and Response to FAS

Recognition and correctly responding to fear, anxiety, and stress (FAS) in animals in a shelter environment is important due to the negative effects FAS has on animal health and behavior. Shelter animals represent a population that are more likely to experience stress due to the many unfamiliar individuals, confined enclosures, and noise in the animal shelter environment. When FAS is recognized and addressed, the animal will be less likely to become reactive or aggressive and will be more cooperative in the long run. This reduces the risk of harm to the animal, shelter staff, and potential adopters. If FAS in animals is disregarded, reactivity and behavior problems will worsen over time and become difficult to correct.

1.1 Recognition of FAS

1.1a Behavioral Cues and Behavioral Scoring Tools

Early recognition of FAS is important to prevent further deterioration and allow for easier intervention. FAS in animals can be recognized via animal behavior or biochemical cues. There are certain behaviors and signs of body language associated with FAS, though different animals express these behaviors to various degrees.

For most animals, signs of FAS include: piloerection, trembling, tucking the tail close to the body, leaning backwards, crouching when approached, flattened ears, holding the head low with eyes averted, and glancing around frequently. Abnormally high/low activity, eating, or drinking also indicate stress. The animal will also freeze or attempt to flee, resorting to aggression when perceiving an inescapable and increasing threat (Moffat 985-986). Animals may display displacement behaviors such as out-of-context yawning, smacking/licking lips, or grooming in order to decrease arousal and cope with stress. These subtle FAS-based behaviors should be monitored in order to avoid escalation into aggression (Moffat 987). Some other FAS behaviors specific to dogs include excessive barking, hypersalivation, nose licking, and inappropriate urination/defecation. Panting and paw-lifting are indicators of short-term stress. When experiencing higher stress, dogs can become more active and startle more. (Hiby et al.

390). A highly alert and rigid dog is ready to display aggressive behavior. Meanwhile, the anxious cat will crouch with ears swiveling sideways, muscles tensed, and tail held close to its body, ready to flee. It will aim to warn others from approaching with its arched back, dilated pupils, curved swishing tail, bared teeth, and claws exposed to swat.

A behavioral assessment upon intake could help in determining the reactivity of the animal and the next course of action for the shelter. This would ensure that more fearful or aggressive animals are handled by experienced staff, preventing risk of injury to inexperienced volunteers or the possibility of sensitizing the animal further. In surrender cases, getting information from the previous owner about the animal's behavioral patterns and history is helpful to understanding potential triggers and recognizing changes that may be health-related.

In cases where no history is available on an animal, behavioral assessment questionnaires can be used to help determine the animal's temperament. The BSAVA Canine/Feline behavioral questionnaire, which has both a canine and feline version, is reliable for assessing the animal's behavior. It is a 10 page, detailed questionnaire originally intended to be used by a veterinarian to help obtain a patient's behavioral and medical information. While this form is too long to be practical in a shelter environment, a shortened version focusing on behavior may be helpful in shelters. Another canine behavioral assessment option is the C-Bar Q. The original long C-BARQ, a 100-item questionnaire, has been used reliably in research and other settings for evaluating dog behavior and temperament but is too long for use in shelters (Wilkins et al. 2). The Short C-BARQ (C-BARQ(S)) was developed by combining questions from the original into 42 questions designed specifically for behavior problem screening of surrendered shelter dogs. It can be completed in less than 10 minutes and found to be comparably valid, since it had high agreement to the original version (Wilkins et al. 5-8).

The Feline Spectrum Assessment by the American Society for the Prevention of Cruelty to Animals (ASPCA) is helpful for evaluating cats' comfort level with humans and their suitability for a home or for return to the field. It is a short 5-minute test, done four times over the first 3 days of intake. Any evaluator with basic cat communication knowledge and mobility can conduct the test, as it includes detailed instructions on use of the tool. The test consists of 4 evaluation items: 1) greeting the cat, 2) cracking the cage door, 3) interaction with the toy, and 4) touching the cat with a wand. The cat's response behavior to each item is checked off on the scoring sheet, and points are totaled to place the cat on a scale from 1 (Extremely Unlikely to be Socialized) to 50 (Extremely Likely to be Socialized). A combination of certain engaged and relaxed behaviors immediately identifies a cat as social and ends the test (ASPCApro "The ASPCA's Feline Spectrum Assessment Training Manual and Guide.").

This test was used to evaluate the sociability of cats post-adoption, where completely social control cats were compared with cats suspected to be less sociable. As a whole, FSA cats showed lower affection and greater fearfulness than control cats. Shelter cats rated "Less Likely to be Sociable" or below had reduced welfare in a home and reduced owner satisfaction, in comparison to cats rated "Likely to be Sociable" or above (J. Ellis et al. 3-6). The success of the Feline Spectrum Assessment in predicting post-adoption sociability here supports its validity.

1.1b Biochemical Cues

One commonly used biochemical method to determine FAS in animals in research is to measure their cortisol levels, as cortisol is a hormone whose release is triggered by stress. Urinary cortisol-to-creatinine (C/C) ratio is a less invasive method to determine the dog's cortisol levels. This is determined by collecting urine and analyzing it in a lab. Cortisol within the body is excreted by the kidneys into urine, with higher blood concentrations of cortisol correlating to higher cortisol concentrations excreted into the urine. As the urine accumulates in the bladder for hours before being excreted, urine cortisol concentrations provide a measurement of the animal's average stress over a period of time compared to a measurement at a single point in time.

In one study where both behavioral and biochemical measurements of stress were taken, increasing C/C ratios were significantly correlated with indicators of stress (more drinking, startling, walking around), and dogs with higher C/C show higher arousal compared to dogs with lower C/C (Hiby et al. 389-390). The fact that behavioral indicators of stress correlate increased C/C ratio validate this biochemical method for detecting stress in dogs.

Although biochemical methods provide a more objective measurement of stress, determining FAS with biochemical methods is inconvenient for a shelter since urine sample collection and lab analysis is required. Realistically, most shelters will be observing behavior.

1.2 The Confounding Factor of Pain (2 pgs)

1.2a Distinguishing Pain

It is difficult to distinguish behaviors caused by FAS from those caused by pain. Physiological signs do not solely indicate one or the other, and each animal has a unique response to pain stimuli. Disease and surgery can also be underestimated as causes of pain. Stress and pain often occur simultaneously, but there are certain ways to help tell them apart.

The American Animal Hospital Association and the American Association of Feline Practitioners (AAHA/AAFP) have developed a set of Pain Management Guidelines published in 2007. It gives four general signs of pain – 1) loss of normal behavior, 2) expression of abnormal behaviors, 3) reaction to touch, 4) physiologic parameters – and lists the specific signs of these (Hellyer et al. 241). Stressful animals are difficult to distract/calm by interaction or handling, administering more opioids does not help them, and a source of pain cannot be readily identified. However, an animal in pain can be temporarily distracted/calmed by interaction/handling, helped by repeated opioid doses, and has an identifiable pain source. Touching the painful or tense area will cause them to look at it (Hellyer et al. 240).

The updated 2022 Pain Management Guidelines adds that chronic pain can be identified by observing abnormal behavior and awkward posturing, as animals move to accommodate the source of their discomfort (Gruen et. al 59-62). Constant observation is difficult for a shelter due to limited resources, but recording any irregularities noticed by staff or volunteers would be helpful to catch signs of pain.

Several pain scales exist to aid with determining the presence and level of pain in a more standardized way. Ideal times to pain score an animal at the shelter would be upon arrival, after

they experience illness or injury, or after any sudden behavioral change. Enough time should be provided for the animal to naturally adjust before testing for each item of the test.

The Glasgow Pain Scales, which has both a canine and feline version, are short and convenient for a busy shelter. It covers both physical and behavioral signs of pain. The Glasgow Composite Measure Pain Scale (CMPS) is a simple 7 question scale developed to assess pain in dogs. The evaluator approaches the kennel, opens the door, encourages the dog to come, walks the dog slowly if possible, and touches the painful area. The dog's response is observed and a qualitative answer on the CMPS is selected for each item. The CMPS has been used in clinical trials, and would be most applicable for animals receiving medical treatment or recently had surgery (Reid et al. "CMPS-SF" 97). One limitation of the CMPS is that it might not be as helpful when the source of pain is unknown.

A short-form CMPS (CMPS-SF) was developed with the help of vet surgeons for quick dog pain scoring in clinics and is used for research as well. The items in the original qualitative scale were shortened and ranked into a numerical scale, but evaluation procedure is similar to the original. When the CMPS-SF was used across three animal hospitals, the median score for dogs considered to need analgesia was 7, 7, and 8 while the median score for those that didn't was 4, 2.5, and 3 (Reid et al. "CMPS-SF" 100). The validity of the CMPS-SF is shown through the significant difference in pain score between dogs considered to need analgesia versus dogs that didn't, as well as through the score consistency across the three veterinary hospitals.

A feline version (CMPS-Feline) has been developed as well. It includes 4 questions on cat vocalizations, body language, behaviors, and ear position and muzzle shape. In a validating study for the CMPS-Feline, 80% of cats were correctly identified as needing analgesia by the test with 17.6 percent misclassification (Reid et al. "Definitive Glasgow acute pain scale for cats" 449); this shows that the scale is relatively accurate.

Another pain scale option for cats is the Feline Grimace Scale (FGS), which provides detailed FGS training resources intended for general use, not just for vets. Scorers evaluate five action units, which are ear position, eye shape, muzzle tension, whisker position, and head position. This test was validated by geometrically measuring distances and angles of cats' ears, face, and head positions from video screenshots (Evangelista et al. 1-3). The FGS has high internal consistency and significantly distinguishes painful cats from non-painful cats (Evangelista et al. 4-7). It only takes 1 hour to film the cat, so pain assessment could become very convenient if machine learning were integrated with the FGS for shelter use.

The advantages of these scales is that they are fast and simple to use in the shelter setting, with minimal training required. Studies that validated Glasgow Pain Scale and Feline Grimace Scale included patients suffering pain from varied sources, which makes the studies more generalizable to different cases of pain (Reid et al "Definitive Glasgow acute pain scale for cats" 449; Evangelista et al. 4). But both these two scales, along with many other available pain scales, were validated by veterinary professionals and intended for use in a clinic. This may undermine its accuracy when used in a shelter environment by non-veterinarians. In a study evaluating Glasgow Pain Scale agreement, it was shown that vet experience is a factor in score accuracy.

Vet professionals had good inter-group agreement in pain scores, while the vet students had poor agreement with each other and tended to score higher than the vet professionals (Marco-Martorell et al. 5). It was shown that gender could play a role in pain scores as well. (Marco-Martorell et al. 10). Therefore, it may be best for shelter staff to evaluate pain in a group for a more consistent pain score and work with a vet professional. There is still a need for pain scores specifically for recognizing pain in shelter settings, which has a higher-stress environment and staff with widely varying levels of experience.

1.2b Managing Pain

Pain management is important as pain decreases quality of life, negatively impacts health through pain-induced stress responses, and affects behavior which risks staff safety. Pain can be considered as adaptive (normal response to tissue damage that responds to treatment, such as inflammation) or maladaptive (physical changes from untreated pain that causes pain to be generated spontaneously, such as nerve damage or abnormal stimulus processing). The longer the pain is left unmanaged, it will more likely turn maladaptive and difficult to treat. It is essential to bring animals suspected of pain to a veterinarian soon and prevent further deterioration.

Local or topical anesthetics help with managing smaller discomforts. Also, preemptively giving pain medication before surgery helps dampen the pain response and reduce doses required to maintain anesthetic (Hellyer et al. 239). Reducing pain in other veterinary procedures would also help, and anxiolytics, sedation, or analgesia should be used when an animal must be handled excessively. Dosage and analgesic drug type vary by affliction and animal. Use of multimodal intervention, which means utilizing multiple pain medications that target different receptors/mechanisms, helps to prevent over-use of one medication (Gruen et al. 56). This reduces side effects and reduces overall pain. If unsure of whether pain medication should be used, it would be best to consult a veterinarian.

Using more careful handling methods which are adjusted to accommodate the pain will help the animal. If the source of pain cannot be resolved, palliative care with pain medication should be implemented to maintain good quality of life and prevent the pain from becoming more maladaptive. Preventative care such as dental maintenance and managing obesity to prevent osteoarthritis reduces pain development later (Hellyer et. al 236). While consistent care for one animal may be an unrealistic goal to achieve for each animal in the shelter environment, other interventions such as environmental accommodation, positive interaction to lower FAS, regular exercise, and diet/weight management to reduce joint stress is more realistically achievable (Hellyer et al. 244; Gruen et al. 70). Cold therapy via applying cold temperatures to painful areas is also an easy way to provide relief, if approved by a veterinarian for that animal.

1.3 Behavior Management

1.3a Prevention of Escalation

Behavior issues relating to FAS often stem from fear and learned reactivity. When faced with a negative outcome or threat, the animal's goal is to avert the threat. They will first try to do

this through appeasement, displaying their discomfort with mild fear behaviors and attempts to flee or freeze. However, their expressions of appeasement are often unnoticed or ignored for the sake of efficiency, and the animal learns to escalate quicker to receive a favorable outcome. To prevent this, staff and volunteers should be trained to identify and respond to subtle FAS signs. The animal should be given the option to flee when uncomfortable, and gentle handling with minimal restraint should be used by staff. Noticing and responding to these signs will allow the staff to de-escalate the situation, preventing the animal from learning aggression (Moffat 984).

Vets help animals acclimate to the exam room by spending time sitting on the floor with the animal and subtly offering food treats (Moffat 988). This approach can also be utilized by shelter staff. Acting in a non-threatening manner by kneeling to the side of the animal or cage door, minimizing noise and speaking softly, and putting smaller animals on one's lap can make them feel more comfortable. Approaching frontally and looking directly at the animal, walking or reaching over them, or cornering them should be avoided.

In general, good behavior must be intermittently and meaningfully rewarded for all dogs to maintain positive emotions and good behavior, especially in the stressful shelter environment. Stressful procedures can be made more pleasant by providing toys and treats, handling gently, and removing other animals. In shelters, it's common to see a mix of defensive aggression motivated by danger of harm and offensive aggression motivated by resource control (ASV 45-46). Since reactivity is mostly learned, it's important to monitor animals that show mild FAS and displacement behaviors to avoid escalation up a "Ladder of Aggression" (Shepherd 14-15).

1.3b Behavior Tracking Methods

It is helpful to track animal behavior patterns during their stay at the shelter in order to recognize changes, which may indicate worsening behavior problems, increased stress, or even medical problems. Grading each animal for friendliness upon intake and documenting their score on their kennel card should be feasible for shelters. A green, yellow, and red color-coded system may help with categorizing animals as all-volunteer, adult-only, or manager-only interactions. High-FAS animals should only be handled by experienced staff, as allowing less experienced caretakers to interact with reactive animals risks injuries and inadvertent reinforcement of bad behavior for the animal. Additionally, keeping a paper or other space near the enclosure where staff and volunteers can write down observations during their interactions with an animal (ex. resource-guarding) is suggested to facilitate identification and communication of the most apparent cases of animals experiencing FAS in the shelter environment. Utilizing kennel cards to record the food intake of animals would help monitor their wellness.

Normal eating and drinking patterns can indicate that an animal has habituated to the new environment. Real-time behavior tracking computer programs like the Observer software package from Noldus Information Technology have been found to be highly accurate for monitoring animal eating and drinking (Eagan et al. 2). Computer programs may become helpful in the future for high volume shelters and overcome a lack of human resources as an easy and cost-effective way to track important animal behaviors, but it is still in development.

1.3c Examples of Behavior Modification Programs

Behavioral modification programs are extremely important in correcting reactivity and aggression. This section will discuss two examples of dog behavior modification programs supported by the ASPCA in order to identify common methods to treat behavior issues.

The first is the Wenatchee Valley Humane Society (WVHS) behavior treatment program. Its staff and volunteer protocol is available for download in an ASPCA article, along with kennel cards and other resources. The second is from an ASPCA observational study on rehabilitation of fearful shelter dogs, abbreviated as “ASPCA’s program/study” from this point on.

Upon intake, dogs’ status is evaluated from their behavior and fear in both programs. The WVHS program rates the dog with the Fear Free 0-5 FAS scale to list them as Available or Unavailable. The dog is also categorized based on its specific behavior issues. Meanwhile, the ASPCA’s program conducted subtests by replicating environments of the home and shelter environment. Staff scored boldness, sociability, and aggression scales, then gave an overall letter grade from A (no fear) to D (severe fear). They were re-evaluated about every 21 days, for a total of 6 tests throughout the program. After intake, both programs provide a safe decompression space for dogs. However, dogs with longer acclimation periods did not have faster behavioral improvement, indicating that behavioral intervention is still needed. (Kristen Collins et al. 7)

Both programs used weekly staff meetings to discuss each dog’s progress throughout their behavior modification program and the next steps to take. The programs utilized desensitizing and counterconditioning techniques for triggers like handling. The ASPCA’s program worked on “Socialization with People, Leash Application and Walking, and Handling” specifically in its behavior modification protocol through 15-minute daily treatments. Reserving toys, high-value foods, and social interaction with other dogs to these sessions helped create a positive association with human interaction and training. A “helper dog” was often implemented in these sessions to encourage interaction with people.

WVHS’s program instead describes how general staff and volunteers can support dogs’ needs. It utilizes regular positive interaction as well, like greeting dogs and putting treats in their kennels in passing. In addition, staff can clicker train dogs to react calmly to stimulus. Behavior protocol for specific types of issues, such as utilizing free feeding to reduce resource guarding, is also listed in the program (WVHS “The Wenatchee Valley Humane Society’s Behavior Modification Program”). WVHS kennel cards lists ways to support its dogs via “tossing treats”, “allowing” the dog to approach, and “respecting” their choices (WVHS “I’m a Polished Pup!”). This is helpful in educating volunteers and adopters, which ensures regular positive interaction with humans and reduces FAS over time.

In the ASPCA’s Program, psychotropic medications were used to assist with treatment by reducing fear. A combination of fluoxetine and gabapentin were most effective and didn’t have many observable side effects. Medication was weaned off over 2-3 weeks before graduation. The initial medication protocol had to be revised when the program moved to North Carolina, as almost every dog needed psychotropic medication 8 weeks into treatment. Use of psychotropic

medications is important for putting animals in a more relaxed state and improving their response to behavior treatment.

The ASPCA's program's graduation rate was 86.17%. Graduating dogs spent around 96 days in the program and averaged 78 treatment sessions. On average, graduating dogs received an initial evaluation score of C which improved to a B by their final evaluation. 99% of graduates were adopted after the program, and most adopters who responded to post-adoption surveys reported high satisfaction. Non-graduating dogs averaged 86 sessions over 125 days, and had lower average initial and final grades compared to graduating dogs (Kristen Collins et al. 6). Although the ASPCA's observational study cannot prove causation, it shows that the majority of fearful dogs enrolled in the behavior modification program showed largely positive outcomes. Programs like this can help guide other shelters to develop and employ their own behavioral modification programs.

These two examples showed how standardized behavioral evaluations, calm decompression space, enrichment, staff communication, regular behavior modification sessions, positive engagement with the dog, and use of prescribed psychotropic medication are important parts of shelter behavior management programs.

Section 2: Environmental factors

The unfamiliar surroundings of the animal shelter are a major source of fear for admitted animals. Measures must be taken to design an environment that is low-stress, allows animals to feel safe, and facilitates habituation to the shelter. A comfortable and spacious enclosure, roommate companionship if applicable, and a predictable schedule with regular enrichment will heighten the animal's quality of life. Pheromone use may also help calm the animal.

2.1 Enclosure and Environment Design Considerations

The design of the facility has an effect on shelter animals' mental states and level of fear, anxiety, and stress. A stable enclosure that is easy to sanitize, accounts for the animal's needs, and has a comfortable ambience will help reduce FAS.

Solid walls and doors are recommended, but wire or slatted floors are unacceptable due to discomfort for the animal. A sealed, impermeable surface like urethane or epoxy resin makes good flooring; tile flooring may be used as well but is harder to clean. The cracks where floor meets wall should be sealed so that pathogens do not collect there. Drainage should be implemented to maintain a comfortable and clean environment. Gently sloped floors that allow wastewater to run into covered drains would help with this. Outdoor enclosures should also have similar designs (ASV 18). Humane temperature conditions for animals should range from 20°C (68°F) to 30°C (86°F) (National Research Council et al. 163), depending on their conditions and age. Partially or fully outdoor-housed animals must protect against weather, temperature, predators, harassment, or escape; if the outdoor temperature falls below 10°C (50°F), dogs should not be kept in permanent outdoor runs (USDA p.164). Additional temperature adjustment measures should be taken if the animal is not comfortable. Humidity should be kept between

30% and 70% (USDA p.40). Proper ventilation to keep good air quality in the whole shelter is essential for well-being of both animals and personnel, and may need to be adjusted seasonally. The air quality at the animal's height should be measured (since ammonia levels can differ at human height versus the animal's height). Ammonia exposure can irritate animals' eyes and nose (Hurt et al. 2).

Light stress can also affect animals at the shelter. Exposure to natural sunlight is essential in following the circadian rhythm and improving well-being of people and animals. Artificial light should be placed away from the rest area when possible and follow the natural light cycle to promote rest (Hurt et al. 3). For example, permanent light fixtures should not shine directly at a bed or in the animal's hiding place, and should be turned off at night.

The facility design should include noise consideration. Noise levels in shelters can get up to 100+ decibels especially in dog kennels, which is harmful to both the shelter's animals and people (Coppola et al. 1). The excess noise causes overstimulation and hearing loss, elevates stress, and drives away potential adopters. Shelters designed with concrete walls and exposed metal roofing increase noise, since these nonporous walls reflect the noise within the room. A design solution is using porous sound-reducing panels or walls to help absorb the sound (Hurt et al. 1). In addition, using apartment-style quarters like closed off rooms with glass windows instead of dog runs or cages may help contain the noise (Coppola et al. 4-5). Halfway-up viewing windows on kennel doors surprise dogs when people walk past, so they should be avoided. Since shelter dogs themselves cause most of the noise, teaching quiet behavior using positive reinforcement, such as feeding toys and puzzles, can help reduce barking (Hurt et al. 1-2; Coppola et al. 4-5).

In their primary enclosure, the animal must have enough room to move normally and engage in natural behavior. Dog enclosure sizes vary by breed, but cats generally require at least 8 feet of floor space and also access to elevated surfaces. Animals must have a soft elevated resting place to keep dry and comfortable, and their eating and sleeping area must be separated from the area they urinate and defecate in. All animals should have the opportunity to hide from view and feel secure. This is especially important to reducing FAS, as stress scores of post-operation cats and cats housed in a shelter significantly decreased when a hiding space was provided (Griffith et al. 1155; Vinke et al. 90). Enclosures with multiple compartments are very beneficial, as they offer the animal more choice for hiding and are easier/safer to spot clean (Wagner et al. 639). The ASV Guidelines include diagrams of proper enclosures for both dogs and cats (ASV 15-16). Consulting a veterinarian and a shelter design architect would be helpful in creating the best shelter environment.

2.2 General Care

In general, animals should have a stable environment with a consistent daily routine and consistent caregivers if possible. When events are more predictable, animals experience less stress as they habituate and know what to expect. Then, positive associations with feeding time and play can form which promote positive states of mind (ASV 45).

Enrichment is a required part of animal care. In multiple shelter studies, it's shown that enrichment reduces undesirable vocalization, reinforces calm behavior, and reduces aggression in dogs. This can come in the form of human interaction; Two 15-minute sessions of calm human accompaniment and interaction daily was enough for a majority of previously fearful dogs to pass an aggression screening test (Willen et al. 60). Twice daily in-kennel training for calm behaviors with food toy enrichment also led to significantly less jumping and vocalizing (Herron et al. 689-691). Simply having passerby toss a treat in kennels helps counter-condition dogs' fear of humans approaching their kennel, resulting in significant decrease in decibels of shelter noise over time. As passerby participation increased, noise levels decreased (Carrero and Bennett 46).

Food enrichment was also found to reduce aggressive vocalization and climbing the pen door post-enrichment compared to a control group (Antonino et al. 31-32). Calming food enrichment in the form of a lickable KONG® toy led to fewest vocalizations during morning kennel cleaning, and so did tactile toys and blankets (Dare and Strasser 8-9). In addition, other forms of enrichment like lavender scent or classical music also help promote relaxed behavior and reduced fear in dogs (Dare and Strasser 9; Willen et al. 59-60).

For cats, toys that promote hunting behavior such as chasing and pouncing are good enrichment. These include fishing rod toys, wire-based toys, balls, or rubber bands on a string. Food puzzle toys are also good mental stimulation. Hiding spaces, like boxes and tunnels, and elevated spaces such as cat trees or shelves enrich the environment as well (S. Ellis 903-905). Some cats are active responders that aim to escape or gain attention, and those benefit from stimulatory enrichment like food toys. Others are passive responders trying to feel more secure, and these benefit more with a hiding place and other physical enrichment (S. Ellis 909).

Playgroups and play behavior are important for good welfare and social contact as well. Dogs should be selected based on health and behavior to maintain safety, and the number of dogs in playgroups depends on the shelter's resources. Dedicated enrichment spaces such as play yards or play rooms should be equipped with double doors and protected from the elements. Intraspecific interaction can also increase cats' welfare as long as they are compatible and resources are adequate (S. Ellis 903).

2.3 Co-Housing

Co-housing animals can be beneficial since it promotes social contact, but the animal's well-being should be prioritized. Trained staff should select animals for co-housing based on sex, behavioral attributes, and other factors. The animals should be first introduced outside the enclosure to minimize risk to and determine if they are a good match. No more than six cats and no more than four dogs should be in a group-housing enclosure together for safety and disease management (ASV 17). Animals should be provided with plenty of space and resources (toys, food and water, hiding spots, perches) to prevent resource guarding behaviors and conflict. Young animals should be housed with their littermates and mother or with similar aged animals to aid their development. Additionally, prey species should still be housed away from predator species since the presence of predator species increases stress.

One study shows the benefits of group housing for reducing stress in dogs with existing social skills. Dogs with adequate dog-dog social skills as determined by the Match Up II test were “single-housed” or “pair-housed”, and pair-housed dogs were matched based on staff information, energy levels, and size. The study found that pair-housed dogs had lower lengths of stay and exhibited less signs of stress such as lip licking, whining, and ears back compared to single-housed dogs. Though the C/C ratio didn’t significantly differ between single and pair-housed dogs, the pair-housed dogs had a greater average decrease in C/C ratio over 7 days (Hecker et al. 11-12). Overall, group housing with an appropriate match seems to decrease FAS behaviors and stress.

2.4 Pheromone Use

Pheromones have been reported to reduce the stress of pets and have been implemented in some shelter settings as well.

A synthetic version of feline facial pheromone (FFP) reduces anxiety in cats in unfamiliar environments. In both sick and healthy cats hospitalized at a clinic, FFP-exposed cats showed significantly more facial rubbing, grooming, and interest in food compared to control cats (Griffith et al. 1155). The FFP-exposed cats also showed more lying and sitting, which indicate calmness, but less sleeping. These results show that FFP may help cats adjust to their environment. However, a study done specifically on shelter-housed cats found no significant difference between a placebo and pheromone diffuser in reducing stress scores in shelter cats or reducing upper respiratory tract infection incidence, which is affected by stress (Chadwin et al. 418). Though pheromone diffusers have been found to be effective in home or clinical settings, cats in shelter settings experience different kinds of stressors and are housed near other highly stressed cats. Use of pheromones may be helpful, but not adequate for the specific needs of shelter cats.

Another synthetic pheromone, the Dog-Appeasing Pheromone (marketed as Adaptil), has been shown to help dogs in the vet clinic and at new homes. In one study, dogs that normally exhibited aggressive behavior at the vet clinic had lower measures of anxiety and higher measures of relaxation when exposed to DAP (Mills et al. 121-122). In another study evaluating newly adopted puppies fitted with a collar containing DAP, a smaller percentage of DAP puppies showed signs of fear on both day 3 and day 15 compared to control. Puppies wearing a DAP collar habituated to their owner and environment faster compared to control puppies as well. The DAP group also did not show as many fearful behaviors when faced with an unfamiliar person or environment compared to the control (Gautier et al. 711). The study concluded that DAP may help shelter dogs react better to unfamiliar situations and get used to their new environment, but this has not been tested in a shelter environment specifically.

Conclusion

This paper discussed methods to recognize/respond to FAS in shelter animals and adjust the environment to reduce FAS. Using behavioral cues and the help of scoring tools, the level of

FAS can be determined and presence of pain can be identified. Behavior traits can be tracked on kennel cards to facilitate communication about any areas of concern. Successful behavioral programs utilize in-kennel enrichment, weekly evaluations from behavioral specialists, daily counterconditioning/desensitization treatment to triggers (including handling, leash walking, socialization with people, and meet-n-greet), and psychotropic medication prescribed by a veterinarian as needed.

The animal's enclosure must have comfortable temperatures (20°C to 30°C), regular ventilation, lighting that follows natural day-night cycle. Tolerable noise levels can be achieved by installing sound absorption panels, using apartment-style housing, providing calming enrichment like food toys, and training dogs to behave calmly. Enclosures should allow for natural behavior, include a dry place to rest, and have separate eating and eliminating areas. A fixed daily schedule, ample enrichment, playgroups, and co-housing all help reduce animal FAS. Pheromone use may be able to improve welfare as well.

By following these recommendations shelters can take actionable steps to reduce FAS in the shelter, will lead to less behavior problems, less shelter backlog resulting in decreased rates of euthanasia, improved animal welfare, and increased adoption rates with higher adopter satisfaction.

Discussion

More research is needed to develop better shelter-specific tools. Biochemical cues have been validated as an indicator of stress in shelter dogs, but are difficult to use in shelter settings and can be confounded by pain. Meanwhile, behavioral evaluations and use of scales may be confounded by bias. Distinguishing pain is still difficult as well, especially internal pain. Many scales for behavior and pain have been validated by veterinary professionals, but the validity differs between veterinary professionals and non-professionals. Therefore, adjustments may need to be made to fit a general shelter environment. Limitations to shelter resources can make it impossible to track and monitor the behavior of individual animals, especially since many would show FAS in the shelter setting. Use of computerized behavior monitoring programs may help with this. In addition, some animals in behavior programs were unable to see improvements. Alternative options for those animals other than barn placement and behavioral euthanasia could be explored in future research. The helpfulness of pheromone use is disputed, but there is a lack of research on the effects of pheromones in a shelter environment.

Acknowledgements

Thanks to Sydney Sill for her support, guidance, and expertise which have helped me throughout the writing of this paper.

Conflict of Interests

The author declares no conflicts of interest related to this work.

Works Cited

- Antonino, Gustavo Vieira, et al. "Effects of Two Types of Environmental Enrichment on the Behavior of Dogs in Shelters." *Journal of Veterinary Behavior*, vol. 80, May 2025, pp. 28–38, <https://doi.org/10.1016/j.jveb.2025.05.004>.
- ASPCApr. "The ASPCA's Feline Spectrum Assessment Training Manual and Guide." ASPCApro.org, 2016, www.aspcapro.org/sites/default/files/ASPCA-FSA-manual-2016.pdf.
- The Association of Shelter Veterinarians. "The Guidelines for Standards of Care in Animal Shelters: Second Edition". *Journal of Shelter Medicine and Community Animal Health*, vol. 1, no. S1, Dec. 2022, pp. 1-76, <https://doi.org/10.56771/ASVguidelines.2022>.
- Carrero, Jamirelis Carrero, and Sara L. Bennett. "Effect of Classical Counterconditioning (Quiet Kennel Exercise) on Barking in Sheltered Dogs." *Journal of Veterinary Behavior*, vol. 77, Sept. 2024, pp. 44-51, <https://doi.org/10.1016/j.jveb.2024.09.004>.
- Chadwin, Robin M., et al. "Effect of a Synthetic Feline Facial Pheromone Product on Stress Scores and Incidence of Upper Respiratory Tract Infection in Shelter Cats." *Journal of the American Veterinary Medical Association*, vol. 251, no. 4, Aug. 2017, pp. 413–20, <https://doi.org/10.2460/javma.251.4.413>.
- Coppola, Crista L., et al. "Noise in the Animal Shelter Environment: Building Design and the Effects of Daily Noise Exposure." *Journal of Applied Animal Welfare Science*, vol. 9, no. 1, Jan. 2006, pp. 1–7, https://doi.org/10.1207/s15327604jaws0901_1.
- Dare, Pamela, and Rosemary Strasser. "Ruff Morning? The Use of Environmental Enrichment during an Acute Stressor in Kenneled Shelter Dogs." *Animals*, vol. 13, no. 9, Jan. 2023, p. 1506, <https://doi.org/10.3390/ani13091506>.
- Eagan, B. H., et al. "Behaviour Real-Time Spatial Tracking Identification (BeRSTID) Used for Cat Behaviour Monitoring in an Animal Shelter." *Scientific Reports*, vol. 12, no. 1, Oct. 2022, p. 12:17585, <https://doi.org/10.1038/s41598-022-22167-3>.
- Ellis, Jacklyn J., et al. "Post-Adoption Behavior and Adopter Satisfaction of Cats across Socialization Likelihoods." *Journal of Shelter Medicine and Community Animal Health*, 2025, pp. 1-9, <https://doi.org/10.56771/jsmcah.v4.116>. <https://jsmcah.org/>.
- Ellis, Sarah LH. "Environmental Enrichment: Practical Strategies for Improving Feline Welfare." *Journal of Feline Medicine and Surgery*, vol. 11, no. 11, Nov. 2009, pp. 901–12, <https://doi.org/10.1016/j.jfms.2009.09.011>.
- Evangelista, Marina C., et al. "Facial Expressions of Pain in Cats: The Development and Validation of a Feline Grimace Scale." *Scientific Reports*, vol. 9, no. 1, Dec. 2019, p. 19128, <https://doi.org/10.1038/s41598-019-55693-8>. [nature](https://doi.org/10.1038/s41598-019-55693-8).
- Gaultier, E., et al. "Efficacy of Dog-Appeasing Pheromone in Reducing Behaviours Associated with Fear of Unfamiliar People and New Surroundings in Newly Adopted Puppies." *Veterinary Record*, vol. 164, no. 23, June 2009, pp. 708–714, <https://doi.org/10.1136/vr.164.23.708>.

- Griffith, Cerissa A., et al. "Effects of a Synthetic Facial Pheromone on Behavior of Cats." *Journal of the American Veterinary Medical Association*, vol. 217, no. 8, Oct. 2000, pp. 1154–1156, <https://doi.org/10.2460/javma.2000.217.1154>.
- Gruen, Margaret E., et al. "2022 AAHA Pain Management Guidelines for Dogs and Cats." *Journal of the American Animal Hospital Association*, vol. 58, no. 2, Feb. 2022, pp. 55–76, <https://doi.org/10.5326/jaaha-ms-7292>.
- Hecker, Grace, et al. "Effects of Single- or Pair-Housing on the Welfare of Shelter Dogs: Behavioral and Physiological Indicators." *PLoS ONE*, vol. 19, no. 6, June 2024, pp. e0301137–37, <https://doi.org/10.1371/journal.pone.0301137>.
- Hellyer, Peter, et al. "AAHA/AAFP pain management guidelines for dogs and cats." *Journal of feline medicine and surgery* vol. 9, no. 6, 2007, pp. 466-80. doi:10.1016/j.jfms.2007.09.001
- Herron, Meghan E., et al. "Effects of Environmental Enrichment on the Behavior of Shelter Dogs." *Journal of the American Veterinary Medical Association*, vol. 244, no. 6, Mar. 2014, pp. 687–92, <https://doi.org/10.2460/javma.244.6.687>.
- Hiby, Elly F., et al. "Behavioural and Physiological Responses of Dogs Entering Re-Homing Kennels." *Physiology & Behavior*, vol. 89, no. 3, Oct. 2006, pp. 385–91, <https://doi.org/10.1016/j.physbeh.2006.07.012>.
- Hurt, Moriah, et al. "Promoting the Welfare of Kenneled Dogs: Environmental Considerations." 2015, extension.purdue.edu/extmedia/VA/VA-3-W.pdf.
- Kristen Collins, et al. "Behavioral Rehabilitation of Extremely Fearful Dogs: Report on the Efficacy of a Treatment Protocol." *Applied Animal Behaviour Science*, vol. 254, 2022, pp. 105689, <https://doi.org/10.1016/j.applanim.2022.105689>.
- Levy, Julie, et al. "Shelter Animals and Behavior Problems." *Ufl.pb.unizin.org*, Module 8, May 2020, ufl.pb.unizin.org/integratingveterinarymedicinewithsheltersystems/chapter/shelter-animals-and-behavior-problems/.
- Marco-Martorell, Mireia, et al. "Agreement of Pain Assessment Using the Short Form of the Canine Glasgow Composite Measure Pain Scale between Veterinary Students, Veterinary Nurses, Veterinary Surgeons, and ECVAA-Diplomates." *Animals*, vol. 14, no. 16, Aug. 2024, p. 2310, <https://doi.org/10.3390/ani14162310>.
- Mills, Daniel Simon, et al. "A Triple Blind Placebo-Controlled Investigation into the Assessment of the Effect of Dog Appeasing Pheromone (DAP) on Anxiety Related Behaviour of Problem Dogs in the Veterinary Clinic." *Applied Animal Behaviour Science*, vol. 98, no. 1-2, June 2006, pp. 114–26, <https://doi.org/10.1016/j.applanim.2005.08.012>.
- Moffat, Kelly. "Addressing Canine and Feline Aggression in the Veterinary Clinic." *Veterinary Clinics of North America: Small Animal Practice*, vol. 38, no. 5, Sept. 2008, pp. 983–1003, <https://doi.org/10.1016/j.cvsm.2008.04.007>.
- National Research Council (Waszyngton). *Ad Hoc Committee On Dog And Cat Nutrition*, et al. *Nutrient Requirements of Dogs and Cats*. The National Academies Press, 2006, p. 163.

- PetMD Editorial, and Jennifer Coates. “How Do Calming Dog and Cat Pheromones Work?”
 Www.petmd.com, 25 May 2018,
www.petmd.com/dog/behavior/how-do-calming-dog-and-cat-pheromones-work.
- Reid, Jacky et al. “Definitive Glasgow acute pain scale for cats: validation and intervention level.” *The Veterinary record*, vol. 180, no. 18, pg. 449, 2017. doi:10.1136/vr.104208
- Reid, Jacky, et al. “Development of the Short-Form Glasgow Composite Measure Pain Scale (CMPS-SF) and Derivation of an Analgesic Intervention Score.” *Animal Welfare*, vol. 16, no. S1, May 2007, pp. 97–104, <https://doi.org/10.1017/s096272860003178x>.
- Shepherd, Kendal, et al. *BSAVA Manual of Canine and Feline Behavioural Medicine*. British Small Animal Veterinary Association, 2009, p. 10.
- USDA. *Animal Welfare Act and Animal Welfare Regulations*. 2023,
www.aphis.usda.gov/media/document/17164/file.
- Vinke, C. M., et al. “Will a Hiding Box Provide Stress Reduction for Shelter Cats?” *Applied Animal Behaviour Science*, vol. 160, Nov. 2014, pp. 86–93,
<https://doi.org/10.1016/j.applanim.2014.09.002>.
- Wagner, Denae, et al. “Shelter Housing for Cats: Principles of Design for Health, Welfare and Rehoming.” *Journal of Feline Medicine and Surgery*, vol. 20, no. 7, July 2018, pp. 635–42, <https://doi.org/10.1177/1098612x18781388>.
- Wenatchee Valley Humane Society. “I’m a Polished Pup!” ASPCA-SBR-Resources, Online Image. Accessed 2025.
- Wenatchee Valley Humane Society. *The Wenatchee Valley Humane Society’s Behavior Modification Program*. Accessed 2025.
- Wilkins, Vanessa, et al. “Validation of the Shortened Version of the Canine Behavioral Assessment and Research Questionnaire (C-BARQ) Using Participants from the Dog Aging Project.” *PLoS ONE*, vol. 19, no. 4, Apr. 2024, pp. e0299973–73,
<https://doi.org/10.1371/journal.pone.0299973>.
- Willen, Regina M., et al. “Enrichment Centered on Human Interaction Moderates Fear-Induced Aggression and Increases Positive Expectancy in Fearful Shelter Dogs.” *Applied Animal Behaviour Science*, vol. 217, Aug. 2019, pp. 57–62,
<https://doi.org/10.1016/j.applanim.2019.05.001>.
- Wolf, Peter J., et al. “Measures of Well-Being in U.S. Animal Shelter Staff during 2023.” *Journal of Shelter Medicine and Community Animal Health*, vol. 3, no. 1, Oct. 2024, pp. 1-13,
<https://doi.org/10.56771/jsmcah.v3.81>.

Mathematical Foundations in Hindu Temple Architecture: An Analysis of Geometrical, Proportional and Astronomical Principles in Structural Design and Decorative Motifs

By Bhadra Ajayan

Abstract

Hindu temple architecture has always been a significant cultural structure that not only portrays religious and scientific knowledge. The aim of this study was to investigate the applications of mathematics in Hindu temple architecture - geometry, fractals, and mathematical patterns - contributing to its astronomical, cosmological, and philosophical importance. A qualitative, interdisciplinary research design - systematic research methods and a single case study- was utilized to analyse these principles of temple architecture.

The literature review was a thematic analysis of academic papers and scholarly articles, while the interview was conducted with an architecture professor. The findings reveal that Hindu temples are an integrated system of mathematical principles that aligns them with astronomy, cosmology, and philosophical beliefs and knowledge. Key architectural elements of Hindu temples follow fractal patterns and symmetrical ratios, which not only create aesthetic flow and rhythm in the temple but also connect it to spiritual and ritual beliefs/practices. Therefore, this study highlights the profound applications of mathematics in Hindu temples and how it interconnects several other aspects to function as a cohesive system. Being aware of the unified principles in temple structures enables a greater appreciation of the heritage and civilization. It gives an insight into the advancements and knowledge acquired by such communities about mathematics, space science, and the world around us. Hence, it reinforces that Hindu temples are not just aesthetically appealing structures, but structures that preserve knowledge and beliefs by unifying aspects of mathematics.

Keywords Geometry, Fractals, Cosmogram, Hindu Temple Architecture, Astronomy

Introduction

Historical context of Hindu temple architecture

For centuries, Hinduism has been a significant religion, ardently followed by thousands in the subcontinent of India. One of the most symbolic representations of the religion is the temple, a sacred structure that was built to hold the deity and become a place of worship. *Bairagya* stated that Hinduism itself was often understood as a part of the Indian civilization rather than a religion; therefore, the temples reflected more than the gods but an ideal way of living and connected with divinity in the earlier civilizations. Moreover, the temple's architecture embodies various cultural and historical pertinence that gives value to each sacred edifice. Hindu temple architecture is not monolithic, as it has evolved throughout the centuries and been modified across various regions - from the *Nagara* (North Indian style) to the *Dravidian* (South Indian style) to the *Verasa* (a hybrid between the two). The evolution underscores how temples

are a record of cultures, civilizations, and beliefs, while their architecture serves as an expression of spiritual devotion. Moreover, the structure and architecture of the temple reveal pertinent information about the sociopolitics and ruling of the temple, which existing research doesn't give sufficient importance to.

Mathematics in Ancient Indian Knowledge Systems

Ancient Indian knowledge systems are emphasized as the essential factor in temple architecture. *Vastu Shastra* and *Shilpa Shatra* are the primary two studies and principles of the frameworks of these architectures. *Shilpa Shastra* is the ancient Indian science of art and sculpture, providing rules for designing temples, statues, carvings, and other sacred structures, linked to geometry and symbolism. By specifying proportions and material, the *Shilpa Shastra* helps the divine sculptures and idols become an accurate representation of the cosmic order and establishes a divine space. *Vastu Shastra* is the guidelines used for the spatial position of architecture.

Kaur et. al(2024) discussed that the Modular system of measurements of the human form, a proportional tool to be used in architecture, was inspired by an aspect of the Vastushastra, known as the *Vastu Purusha Madalam*. *Vastu Purusha Madalam*, which translates to “a grid for architecture, reflecting a cosmic being”. The science of *Vastu* is known to comprise 5 elements - water, air, fire, space, and land - and the balance of these elements is the sole reason that Earth can sustain human life, unlike other planets (*Patel, et al.,2023*). The Vastupurushamandal is known as a field of energy present in each building, which protects the inhabitants of the structure. With the head of the vastupurush facing northeast and his feet facing southwest, the grid splits the body into equal squares, with each area being assigned to sanctum, halls, or shrines of the temple.

In addition to this, temple architecture also drew on geometry and ratios, as they were believed to achieve harmony throughout the structure. The Golden ratio(ϕ) is one of the most commonly applied formats in Hindu temples, as this special number, approximately equal to 1.618, is found throughout nature. It is achieved when a line is divided into 2 (a longer part and a shorter part), and the longer part divided by the shorter part must be equal to the entire length divided by the longer part. The application of the golden ratio links the temple to the natural order of the universe, presenting it as a microcosm of the universe.

While the golden ratio and *Vastu Pursha Mandalam* are some of the primary methods and ratios implemented, other constructive techniques and patterns also play a critical role in the architecture of temples.

Mathematics in ornamental motifs

The mathematical applications of a temple extend further from just the structural and spatial aspects, but also to decorative features. *Dokras* describes how Hindu temples contain fractal structures, which are infinitely repeating patterns that iterate at a growing scale of magnification. These fractal patterns are a symbol of “the eternal” in Hindu cosmology. For

example, *Dokras(2020)* outlines how the Virupaksha Temple in Hampi showcases this concept evidently, as the *vimana* of the temple (the pyramidal-shaped tower on top of the temple) has repeating facts that decrease in size and number as the height increases. In addition, tessellations are also a significant aspect, as they are tiles of repeating shapes without any gaps or overlaps. These contemporary beliefs of repetition in patterns reveal a distribution of galaxies following the fractal scales that mirrors the Ancient Hindu beliefs, reflected in the temples, of how *tanmatras* (atoms and molecules) enable the cosmos to replicate itself infinitely. This thereby embodies the eternality of the divine and the boundless recurring cycle of our universe. Yet, while the application of geometry and fractals has been widely explored, the applications of ornamentation motifs and repeating designs/carvings on the Hindu temples haven't been profoundly researched.

Astronomical Alignments and Cosmological Significance

Temples were created to be a representation of the divine order and a microcosm of our universe. Hence, astrology and cosmology play a significant role, as many temples are aligned in the cardinal direction, where the sun rises on the equinox (2 days a year where the sun crosses the celestial equator) and solstice (the longest and shortest day of the year) to portray a direct connection between human activities and cosmic power. While studies have been present around Vedic astrology (known as *Jyotisha*) and such symbolic associations, the direct and profound ways in which astrology influences the Hindu temple structure remain to be a less explored domain.

Hence, this research paper will explore the mathematical foundations in Hindu temple architecture, analysing geometrical, proportional, and astronomical principles in structural design and decorative motifs. By addressing the research gaps, this study will present Hindu temples as more than sacred spaces, but as deliberate representations of ancient mathematical reasoning and cosmological order, looking into many of the aspects and patterns that haven't been extensively investigated. Exploring this integration allows for gaining key insights into how profound ancient Indian knowledge systems were in the construction of Hindu temples.

Methodology

This study adopts a qualitative and interdisciplinary research design, combining systematic research methods with a single case study. The research paper is interpretive in nature, examining the mathematical principles embedded in temple architecture, which are linked to elements of geometry, proportion, and astronomy. The integration of primary and secondary sources is critical in exploring a triangulation of aspects - theoretical frameworks, existing research studies, and lived architectural practices. A literature review was performed utilizing journal articles and authoritative scholarly texts. The selected literature was analysed thematically to identify recurring patterns of principles, the strengths, and gaps in the existing interpretation. The existing scholarship was classified into the following themes: the diverse temple architectural styles, fractals and geometry, sacred proportions, and the astronomical and

philosophical significance of the temple structure and motifs. To complement the secondary research, a single semi-structured interview was also conducted as a case study with a practicing scholar, an architecture professor of Thyagarajar College. This interview primarily delved into the mathematical application in temples (fractals and patterns), with its significance justified through astronomical and philosophical references. The case study was critical in reinforcing the themes analyzed in the literature review and providing unique insights and perspectives into this topic. Overall, this structure effectively combined various research methods to create a cohesive study that investigates the foundations of Hindu temple architecture within the scope of the research.

Discussion

A. Literature Temple

Architectural Styles

The current scholarship uncovers the adaptability and distinctiveness of Hindu temple architecture across various regions of India. It divides temple architecture into 3 main styles - *Dravidian*, *Nagara*, and *Vesara* - with each containing unique designs, patterns, and structures. Sompur & Udachan (2024) focus primarily on comparing the iconography to prove and the symbolism of each of the 3 temple styles, by comparing the tala-mana proportions of the deity statues in 20 temples (quantified data). The analysis of the sculptural plaques strengthens the study's argument that, although the motifs vary regionally, they still refer to the same cosmos and the ashtadikpalas (main deities who guard the 8 cardinal directions). Similarly, Gurme & Patil (2022) discuss these three types of temple architecture, along with Kerala roofs and Bengal terracotta, providing examples in a broadly chronological review format. The author emphasizes the most pertinent features of each of the temple styles (like curved towers for *Nagara* and pyramid roofs for *Dravida* temples). Modi (2024) also explores the 5 Hindu temple architecture styles, like Gurme & Patil(2022), but provides additional insight on materials adaptation in the South and North of India, and the exchange of cosmology knowledge during the integration of dynasties. On the contrary, Singh et. al (2020) shifts away from a standard comparison of the types of architecture to how politics and hierarchies have influenced the temples' style. By studying how the movement across each temple - *Nagara*, *Dravida*, and terracotta - is laid out, it reveals how the outer parakas of the temple accommodate a large crowd while the *Garbhagriha* is designed for an intimate devotion, connoting the Hindu ideology of *samsara* (endless reincarnation cycles) and *moksha* (eternal bliss).

From a holistic perspective, this theme provides a comprehensive analysis of the various types of Indian Hindu temple architecture, case-studying over 20 temples. Moreover, the overall literature contains studies that concentrate on a wide range of tools and data forms (both qualitative and quantitative), reducing the mundanity of effectively exploring various aspects of the theme. However, the overemphasis on *Nagara* and *Dravida* temples is prevalent in the

literature, with styles like Bengal terracotta viewed as peripheral. Hence, a balanced investigation of all temple types would provide a more holistic understanding of the key features and diversity of Hindu temple architecture. Generally, this theme is extremely significant, as it provides a foundational knowledge of the main forms of architecture in Hindu temples, underscoring that the presentation of knowledge and beliefs differs. It also directly links to the study of the Hindu temples as the style of temple contributes to the intentionality behind that choice, which influences mathematical propositions, symbolic meaning, and, most importantly, the devotee's experience.

Fractals and Geometry

The prevailing theoretical framework suggests a recurring pattern of fractals and recursive geometry in the Hindu temple architecture. Fractals are self-repeating patterns across various scales, derived from the Latin term "*Fractus*" or broken. Misra & David (2021) state how the application of fractals in architecture, particularly Indian temples, exhibits the complexity of the pattern and the knowledge-intensive planning that the Hindu temples relied on. They further argue how these fractals parallel the flow of energy across the universe. Although the authors don't explicitly refer to the laws of conservation of energy, their view on the universal energy flow can be interpreted as aligning with the idea that energy undergoes continuous transformation (similar to the perpetual fractals) rather than being created or destroyed. On the other hand, Halai (2019) highlights the fractal geometry as a reflection of ancient metaphysics, connecting it to the holonomy - the ideology where every fragment of the cosmos contains an essence of its whole. Unlike Misra & David's (2019) outlook, Halai (2019) utilizes examples of recursive geometry, like Kierpiński's gasket in the *shikar* of the Kandariya Mahadev temple, to cement its significance. On the contrary, Chathervedi(2022) also explores the visual appeal: how high fractal dimensions (more complex fractals) and symmetry enhance aesthetics and align with innate human preferences. This shows that the application of fractals requires closer cognitive attention from the observer of the structure, increasing the quality of the overall experience. Overall, there is an evident articulation of the pertinence and symbolism of fractals, providing examples of the application in a variety of Hindu temples. However, one major limitation of the existing studies is the lack of quantitative validation for the fractals. Moreover, many research studies fail to address temples that don't have fractals strongly embedded in their structure, which leaves questions of possibilities of alternative designs or reasons why the fractals are absent. This theme directly links to the research question as it shows how the incorporation isn't just a superficial decorative element of Hindu temple architecture, but it exhibits the ideas of temples reflecting a perpetual structure and the order of the universe.

Sacred Proportions

The prevailing theoretical framework suggests that sacred proportions like the Hindu temple designs are primarily determined by certain numerical rules/guidelines and proportions that link the architecture to human and cosmic scale. Dokras(2022) explores a prominent link

between the human body and the temple, as revealed by the ancient literature of agamas. Misra delves into how the proportional grids, like the *Vastupurushamadalam*, were developed in order to establish the temple as a microcosm of the universe, where the squares represent divinity and perfection, while the circles depict human life. This paper presents the temple as a human body, aspects like the layout and divisions according to the Vedic proportion principles. Similarly, Dokras & Jain(1989/2020) also explore the application of *Vastu Shastra*, which is considered to be a set of guidelines that help design various types of temples, and madalam grids like the *Vastapurushamandalam*. Dokras & Jain(1989/2020) strongly emphasize the mathematical background of these shastras, applying and explaining models like Menger's Sponge, Van-Koch snowflake, and Sierpinski Triangle. The pertinence of these self-repeating proportions is also revealed, as the Hindus believe that the universe goes through a perpetual process of being “created, destroyed, and recreated”, portraying the cosmos as being holonomic in nature; therefore, the temple layout also follows the same pattern, making it a microcosm of the universe. On the contrary, Singh(2018) argues that the *Sri Yantra*, a tool that represents cosmic energy, spiritual growth, and symmetry, is derived from the golden ratio and is a more precise design layout for temples. It is argued that although more complex than the common *Vastupurushamandalam*, it has more scope and potential in adequately constructing the plan of a temple, urging more research into this tool’s utility and parameters in Hindu temple architecture. However, the limitation is in the uneven research conducted for temples’ measurements and proportions. While some popular temples like the Kandariy Mahadev temple and the Lakshmana temple are heavily and more frequently researched on with sufficient quantitative data, rudimentary claims are made on most of the other temples. Moreover, the limited research and comprehension of a variety of mathematical proportions, apart from the conventional Golden ratio and *Vastupurushamandalam*, reveal the scope that remains in this area of knowledge. This theme is directly correlated with the research question as it portrays how proportional and quantitative guidelines are significant mathematical aspects in ancient Hindu temple architecture, which connects the earthly and divine beings.

Astronomical significance of Architecture

The current scholarship proposes that the Hindu temples were designed with strong astronomical intentions, where the orientation and proportions correspond to celestial patterns. Shirodkar(2024) underscores the astronomical significance of details and alignments of Hindu temples by exploring the Konark Sun temple and the Sun temple in Modhera. These temples are strategically located and designed in terms of solar alignment, where the rising sun first illuminates the Sanctum and where the sun enlightens the apex of the temple without casting a shadow. The scholar focuses on aspects such as the incorporation of celestial numbers, such as 108 (a number that defines many relationships between celestial bodies), the 24 wheels in the Konark temple chariot that are a representation of *Kala* time. This reveals that the integration of astronomy into temple architecture served a symbolic and ritual purpose, connecting the worship of humans with the cosmic order. On the other hand, Shylaja(2017) explores how temples serve

as observational instruments to better comprehend cosmology - the interplay between light, shadow, and orientation indicates solstice, equinox, and other main celestial days. Even the layout and structure of these temples are sacred numbers like the lunar mansion of the days in a year, which transforms the Hindu temple into a cosmogram. By doing so, it preserves the practical astronomy employed by scholars centuries ago. Overall, this theme has a profound explanation and analysis of how and why certain astronomical aspects are included in Hindu temples, revealing the consistent design logic instead of mere coincidence. This is further bolstered through the examples of temples that the scholars utilize to analyze the astronomical properties, measurements, and significance. On the contrary, a major drawback is that the current available research is primarily centered on more elite, iconic temples, creating the risk of overgeneralization due to the limited evidence for a wider variety of temples. Moreover, scholarship very rarely examines how certain temple features have evolved throughout the years due to the discoveries in astronomy. The historic evolution of these temples due to astronomy provides a clear lens of understanding the period's influence, leaving a gap in the scholarship. For this research, architecture is a key yet sometimes overlooked aspect of temple architecture that reveals the interdisciplinary topics that Hindu temple design involves. It conveys that temples are created not just as an abode for worship, but to preserve astronomical and cosmological knowledge- the integration of science, mathematics, and spirituality.

Philosophical Significance of the Architecture

The existing academic literature showcases that the Hindu temple architecture holds deep philosophical depth, paralleling the beliefs and knowledge of the communities in that time period. Dutta & Adane(2014) focuses on how the application of fractal geometry communicates the metaphysical ideologies, which provoke the visitors in a visual and intellectual manner. The parallel of temples to cosmic unity, interconnectedness to nature, and the human body is also explored in depth, highlighting the introspective significance of the Hindu temple architecture. On the other hand, Ahmad et .al(2019) explores elemental symbolism and cosmological grids rather than prioritizing fractal-geometric symbolism. It goes into the philosophical and ritual aspects of the Hindu culture that the temples accommodate through their various parts, like the Garbhagriha and pillars. It focuses on identifying the various characteristics in temples - implementation of cosmology theory, human measurement theory, and more. In addition, the author depicts how these structures represented the evolving civilizations of Buddhism and Hinduism. Overall, this theme showcases the interiority that goes behind temple designs, covering a variety of different aspects such as the symbolic significance of mathematical grids, designs, and part of a Hindu temple. However, this philosophical aspect should have been explored in more depth, particularly with regard to how the significance of certain shapes, designs, or grids evolved as the civilization and knowledge of the communities did. Knowledge and beliefs are constantly evolving, and understanding how that is reflected in the architecture would have provided a stronger, more dynamic relationship between metaphysical ideologies and tangible structures. Finally, the current scholarship is greatly enriched by this theme of the

philosophical depths of Hindu temple architecture, as it provides a secure foundation to understand the design of Hindu temples. The significance and philosophical understanding of the architectural choices tie together mathematics, cosmology, astronomy, and other factors. This integration, therefore, enables us to perceive how a physical structure is transformed into a sacred space, containing knowledge, beliefs, and cultural value.

B. Case Study

Context

This case study was a semi-formally structured interview to gather first-hand insights from an Architecture professor. The primary objective of this interview was to understand how experts interpret architectural and mathematical aspects of Hindu temple design, and to gain insights that bolster the available scholarship when analyzing the geometry, cultural symbolism, astronomy, and the evolution of the divine space. This interview mainly consisted of open-ended discussions that not only provided a profound understanding of each of the main themes explored in this study but also brought up unique insights and perspectives on how certain elements are inevitably interrelated in Hindu temple designs.

Background

Participant X identifies himself as an Architecture Professor and Dean of Thiagarajar College of Engineering, Madurai, India. Participant X has earned his Bachelor's in Architecture at Thiagarajar College of Engineering(Madurai), Master's in Architecture at School of Planning and Architecture (New Delhi) and PhD at National Institute of Technology(Tiruchirappalli). After completing his studies, he returned to TCE to join the faculty for over 22 years. Some of his publications and achievements include being the Investigative Nodal officer of the “Jal Darohar: Teppakulam and Saravana Poigai Tank Rejuvenation” project, the co-author of the “Investigating the Impact of Urbanization on River Girudhumal” Journal, and being the Coordinator of the International Conference for Urbanization, Ecology and Community Development.

Interpretation

Participant X emphasized how temples integrate mathematical concepts with astronomy, philosophy, and civilizations by stating that a Hindu temple “acts as a single unity” and they “never work in isolation”. He then continued to buttress that point by explaining how every element is interdependent, focusing primarily on the mathematical aspects and their reflection of nature and the cosmos. For instance, his insight into how “most temple complexes...follow symmetrical rules” as “anything in nature is in bilateral symmetry” supports the pertinence of mathematical concepts being employed to reflect natural order in temple architecture. Furthermore, his reference to the statement “you cannot limit God within a boundary”

powerfully encompasses the infinite nature of fractals, cosmos, and faith, showcasing how these temples are intentionally designed with these critical elements.

The interview strongly reinforced many of the literature themes that were previously explored. In particular, Participant X refers to the manipulation of fractals and forms, which is applied to temples to achieve various visual and philosophical impacts. He emphasized design strategies like splitting, where design forms are divided into symmetrical halves to create a center space to house a deity - he compared this process, and its purpose, with nuclear fission (the splitting of atoms), connoting the release of energy. Moreover, he also focused on astronomical alignments and axis mundi in temple architecture. Overall, Participant X's comments strongly supported the theme of philosophical significance by linking the temple architecture with 5 elements, animals, and civilizations, stating that certain temple architectural elements are a result of the interactions between communities and geographic regions.

The interview provides profound insights that reveal the integration of various aspects, including mathematics, cosmology, philosophy, and cultural symbolism into temple architecture. From the fertility and divinity of the land on which temples are constructed to the cardinal directions and the application of geometry, temples are a reflection of the knowledge accumulated by communities. Comprehending how each of these aspects is interdependent in the architecture of a temple allows one to appreciate the beauty, intention, and purpose of the structure.

Participant X emphasized the application of various mathematical concepts and patterns, such as fractals, progressive multiplication, and expanding repetition in temple architecture. These design features create visual rhythms in the temple, highlighting certain key elements and carrying spiritual significance. Furthermore, delving into the purpose of the various parts of the temple (the *garbagraha*, *vimana*, and *mandapas*) reveals how they work together to create a divine, spiritual experience for the devotee. Connecting these parts of the temple with astronomical aspects like the axial mundi, planetary ratios, and cosmogram establishes the temple as a link between the earth and celestial beings.

Moreover, Participant X's perspective on how nature shapes numerous symbols and designs in the Hindu temple is a unique viewpoint that wasn't prevalent in the existing literature. Animal sculptures and engraving are commonly found in temples, as some species, like lizards, mice, and cows, are more evolutionarily advanced than they are perceived to be. Similarly, the presence of the 5 elements (fire, water, wind, land, and space) is critical for temples, as they are the fundamental necessities of life. These aspects suggest that temples embody knowledge and represent the interconnectedness of life through merging various concepts. Implicitly, these components allow the devotees to comprehend their devotion, spirituality, and form as a cohesive experience inside a temple.

Limitations

The study has several limitations that should be acknowledged. Firstly, it heavily relies on secondary literature - historical journals and scholarly articles - which indicates that the

interpretations are constrained by the scope available. Although this provides a strong foundational and theoretical knowledge of the temple architecture, it is limited to the existing research. Moreover, the primary research of this study consisted only of a single interview with an expert in the field of architecture. This limits the potential of various unique perspectives and experiences from practicing scholars that could have been derived from the study to make it well-rounded across all themes. Interviews with a wider range of experts across different fields - architecture, mathematics, art, vastu specialists, and astronomy - would have broadened the insights and reduced potential bias of the case study. Finally, onsite temple architecture data of the measurements, shapes, and sizes were not incorporated in this study. The addition of this data would have significantly bolstered the claims on mathematical ratios and geometry, empirically verifying the principles that were derived from the secondary and primary research. This also doesn't allow for systematic analysis of mathematical ratios and patterns to assess whether the findings can be generalized for most temples across the different regions. However, despite these limitations, this research provides a coherent interpretation for comprehending the foundations of Hindu temple architecture. By exploring the cohesive integration of mathematics, astronomy, and philosophy, this research paper successfully analyzes the pertinence and interconnectedness of these elements in creating a temple.

Conclusion

In conclusion, this research paper has effectively examined the integration of mathematical foundations in Hindu temple architecture through geometry, proportions, cosmology, astronomy, and philosophy. It utilized a variety of research methods to achieve the aim of the investigation, including the literature review and the interview. The literature review, which was analyzed thematically, underscores the role that each principal plays in temple architecture and its symbolic purpose. It collected a wide range of information on the various aspects, establishing that mathematics functions as a medium to create rhythm, movement, and meaning in the temple structures. On the other hand, the interview highlights the critical role that the integration of mathematical patterns, geometry, and fractals plays in reflecting nature, astronomy, cosmology, and other philosophical principles. Participant X provided a more detailed understanding of various mathematical patterns, analyzing different parts of the temple designs to show the applications and its symbolism. This established a more defined correlation between specific designs to elements of nature, cosmos, and ritual practices of Hinduism. These findings from both research methods are pertinent as they reveal that Hindu temples aren't merely a place of worship for a certain group of people that is designed based on the aesthetic, decorative appeal. On the contrary, this study portrays temples as a system that integrates critical mathematical aspects, astronomy, and philosophy to represent knowledge, civilization, and spirituality. Ultimately, it bolsters the idea that numerous principles are unified to create a Hindu temple.

Works Cited

1. Ahmad, A. B., Mujani, W. K., Salleh, M., & Othman, R. (2019). Concepts and philosophy of Indian architecture. *International Journal of Civil Engineering and Technology (IJCIET)*, 10(1), 364–371.
2. Bairagya, S., & Das, M. Tracing the genesis and reflection of heritage temples in India.
3. Chaturvedi, M. (2022, September 7). Fractal geometry and Hindu temple architecture. Viewpoints Which Matter. <https://chaturvedimayank.wordpress.com/2022/09/07/fractal-geometry-and-hindu-temple-architecture/>
4. Dokras, D. U. (2022). Hindu temple architecture. INAC.
5. Dokras, U. (2020). The geometry of Hindu temples. Indo Nordic Author's Collective.
6. Dokras, U. A mathematical temple.
7. Dutta, T., & Adane, V. S. (2014). Symbolism in Hindu temple architecture and fractal geometry – “Thought behind form”. *International Journal of Science and Research (IJSR)*, 3(12), 489–497.
8. Gurme, P., & Patil, U. (2017). A review study on architecture of Hindu temples. *International Journal for Research and Development in Technology*, 8(4), 2349–3585.
9. Halai, C. (2019, October 22). Sacred geometry of Hindu temples. *Indica Today*. <https://www.indica.today/quick-reads/sacred-geometry-hindu-temples/>
10. Kaur, J., Verma, N., & Reddy, T. B. C. (2024). History and theory of design in traditional temple architecture of India. *International Journal for Multidisciplinary Research*, 6(3), 1–13.
11. Misra, S., & David, S. (2021). Significance of mathematics and geometry – formation of temples in India: The golden era of evolution. *International Journal of Engineering Research Updates*, 1(1), 50–55. <https://doi.org/10.53430/ijeru.2021.1.1.0056>
12. Modi, R. P. (2024, January 3). Hindu temple architecture: 5 regional styles across India. *The Architects Diary*. <https://thearchitectsdiary.com/hindu-temple-architecture-5-regional-styles-across-india/>
13. Patel, P. D., & Patel, P. J. (2023). State of art on Vastu Shastra. *Journal of Harbin Engineering University*, 44(7).
14. Shah, K. The geometry of Hindu temples.
15. Shirodkar, K. K. P. (2024). A study of ancient Indian astronomy through architectural marvels. *International Journal of Emerging Research in Engineering, Science, and Management*, 3(2), 06–10.
16. Shylaja, B. S. (2017, December). Astronomical aspects associated with temples. ResearchGate. https://www.researchgate.net/publication/321949369_Astronomical_aspects_associated_with_temples

17. Singh, A. K., Das, V. M., & Garg, Y. K. (2020). Cultural continuity through analysing architecture of Hindu temples in India. *International Journal on Emerging Technologies*, 11(3), 1072–1081.
18. Singh, M. R., Sharma, A., & Chani, P. S. (2018). Proportions in architecture. *International Journal of Research-GRANTHAALAYAH*, 6(3), 126–133.
19. Sompur, S. S., & Udachan, V. (2024). Iconography and symbolism in Indian temple architecture. *International Journal of Engineering Research & Technology (IJERT)*, 13(9).

The Effects of Music Therapy on Cognitive Function in Alzheimer’s Disease By Jenna Kim

Abstract

Alzheimer’s disease (AD) is a brain disorder that impairs quality of life, behavior, and cognitive functioning. Current pharmacological treatments provide limited symptomatic benefit, leading to a growing interest in non-pharmacological interventions, including music therapy, which engages relatively preserved neural systems as well as cognitive and social processes. Across studies, music therapy shows variable cognitive benefits, with more consistent improvements observed in active music therapy approaches. However, findings remain heterogeneous due to differences in intervention formats, duration, outcome measures, and disease severity. While music therapy has the potential to support cognitive function in AD patients, more standardized, rigorous, and longer-term research would be needed to clarify its optimal intervention methods and long-term effects.

Introduction

Every 3 seconds, someone in the world develops dementia—an alarming reminder of the growing global burden of Alzheimer’s disease (AD), which accounts for approximately 60–80% of all dementia cases. Affecting over 55 million people worldwide and projected to reach about 139 million by 2050, AD is a progressive neurodegenerative disorder that currently has no cure (Alzheimer’s Disease International et al., 2024). After the onset of AD, individuals typically experience gradual cognitive decline, characterized by worsening memory, judgment, language, and orientation, often accompanied by psychological symptoms such as agitation, apathy, anxiety, and depression. As the disease progresses, individuals lose their ability to communicate and function independently, increasing their care needs and the burden on caregivers, families, and society.

Although the causes of AD are complex, it is characterized by abnormal amyloid- β deposition and tau neurofibrillary tangles, along with neurotransmitter deficits. Several treatments have been developed to combat AD, but their benefits remain limited. Pharmacologically, cholinesterase inhibitors and the NMDA receptor antagonist memantine provide modest symptomatic benefit for selected patients, and anti-amyloid antibodies can slow decline in early AD (National Institute on Aging, 2024). However, these therapies do not halt or reverse neurodegeneration. Their long-term effectiveness is limited, and they require careful monitoring. Consequently, there is sustained interest in non-pharmacological interventions, including art-based or music-based therapies, physical activity, and meditation, to support cognition, mood, and quality of life.

Among non-pharmacological options, music therapy has drawn particular attention. Music therapy is the clinical and evidence-based use of music interventions to accomplish a patient’s therapeutic goals, provided by trained professionals (American Music Therapy Association, 2010). In dementia care, it encompasses a variety of musical activities, including singing songs together, rhythmic activities, playing instruments, and listening to live or recorded

music. Music can be a compelling tool in AD because it engages attention, timing, and motivation through rhythm and coordinated action, and also social interaction in groups (Peck et al., 2016). In addition, research suggests that aspects of musical memory and related processing often remain relatively preserved even in later stages of AD, likely due to the involvement of brain regions such as the medial prefrontal cortex. This preservation is believed to allow music to evoke emotional responses and autobiographical recall (Jacobsen et al., 2015). Together, these features explain why music-based approaches are being explored to support cognition, mood, function, and quality of life in AD patients, despite varying results across studies in terms of format, dose, and outcome measures.

This review focuses specifically on the effects of music therapy on cognitive function in AD. Most previous studies of music therapy in AD have focused on emotional and behavioral symptoms. However, among the many areas affected by AD, cognitive decline is the central symptom, as it directly contributes to loss of independence and reduced quality of life. Evidence regarding the cognitive effects of music therapy is more mixed, and it is less clear which types of music-based interventions are most effective for supporting memory and thinking in AD. The goal of this review is to summarize current evidence on music therapy for cognitive function in AD and to explore how different music therapy approaches may influence these outcomes. This paper takes a narrative review approach, providing a qualitative summary and interpretation of key studies rather than a formal meta-analysis.

Definitions and Scope

Participants

This review focuses on clinically diagnosed Alzheimer's disease. Some of the studies discussed in this paper include broader dementia samples or people with mild cognitive impairment (MCI). These studies are included when they contain participants with AD or when the findings are likely to generalize to AD.

Cognitive function and measures

In this paper, "cognitive function" refers to both overall cognitive status and specific domains such as memory, attention, language, executive function, and visuospatial abilities. Studies commonly use brief global cognitive screening tools such as the Mini-Mental State Examination (MMSE) or the Montreal Cognitive Assessment (MoCA), and sometimes include more detailed tests that focus on specific domains like memory or executive function (Bleibel et al., 2023). Both global cognitive scores and domain-specific measures are considered when evaluating the potential impact of music therapy on cognition in AD. In addition to cognitive function, some studies also assess neuropsychiatric or behavioral symptoms using tools such as the Neuropsychiatric Inventory (NPI).

Study designs considered

This paper includes several types of studies that examine music therapy in AD, including randomized controlled trials (RCTs) and non-randomized or pre-post study designs. In RCTs, participants are randomly assigned to either a music therapy group or a control group, which helps reduce bias. In non-randomized or pre-post designs, participants are observed over time without random assignment to different conditions.

Active vs. Receptive Music Therapy

Music therapy can generally be divided into two subtypes: active and receptive, which engage patients in different ways. Active music therapy involves direct participation, such as singing, clapping to rhythms, playing simple instruments, and combining music with movement. Receptive music therapy (listening-only; sometimes called passive music therapy) focuses on guided listening to recorded or live music. Both formats have been studied in AD and broader dementia samples to support cognition and quality of life.

Across studies, music therapy shows signs of cognitive benefit, although the effects are inconsistent. Active music therapy, especially approaches based on rhythm or singing, more often reports improvements in global or domain-specific cognition (Bleibel et al., 2023; Ting et al., 2024). Receptive music therapy tends to show more consistent effects on mood and behavior (Tsoi et al., 2018). It also shows occasional cognitive gains, particularly when preferred or familiar music is used with sufficient exposure (Särkämö et al., 2014; Pérez-Ros et al., 2019). Large systematic reviews likewise emphasize variability and generally small or uncertain cognitive effects when findings are combined across diverse music therapy protocols (van der Steen et al., 2025).

Effects of active music therapy on cognition

Research indicates that active music therapy can lead to measurable improvements in cognitive function in individuals with AD. Research studies that combine singing with rhythmic accompaniment or movement often report improvements in global cognition or specific areas such as memory, orientation, and verbal fluency. For example, Wang et al. (2017) found that patients with mild AD who received three daily sing-along sessions over three months (plus standard medication) showed greater improvements in cognition (MMSE, MoCA) and behavior (NPI) than those receiving medication alone. Similarly, Lyu et al. (2018) reported that group singing with familiar or preferred songs, conducted twice daily for three months, improved verbal fluency in patients with mild to severe AD. This improvement was greater than that observed in patients who only participated in lyric reading without melody or who did not receive any intervention. The authors also found that music therapy improved memory and language ability only in patients with mild AD, while it helped reduce psychiatric symptoms and caregiver distress in those with moderate or severe disease.

Studies using rhythmic and movement-based activities have also shown beneficial effects. For example, a six-week music therapy program that combined singing, rhythmic activities, and movement improved cognition, particularly memory, orientation, and language, as

well as mood in people with mild to moderate AD (Gómez Gallego & Gómez García, 2017). However, the lack of a control group limits how strongly these results can be interpreted. Another study involving individuals with moderate dementia found that a six-week music-with-movement program, delivered twice a week, led to greater cognitive improvement than other programs that involved only passive or social activities (Cheung et al., 2018).

Effects of receptive music therapy on cognition

Listening-based receptive music therapy also shows cognitive benefits in AD, though typically less consistent or pronounced than active approaches. In Särkämö et al. (2014), patients with mild to moderate dementia receiving either home-based music listening or singing (guided by caregivers such as family or nurses), combined with weekly group music therapy sessions, improved in both global and domain-specific cognitive functions compared to usual care. Music listening was particularly linked to positive effects on mood and quality of life, though singing provided extra improvements in short-term/working memory and caregiver wellbeing. In another study conducted in a nursing home, residents with dementia who participated in a group program, listening to preferred music five days per week for eight weeks, showed stabilized global cognition (MMSE) and improved functional and emotional outcomes compared with usual care (Pérez-Ros et al., 2019). The program also allowed optional spontaneous participation such as singing, dancing, and clapping.

However, relatively few studies have used purely receptive formats that rely only on listening to music when examining changes in cognitive function in people with AD. Most studies labeled as “receptive music therapy” actually include some degree of active components (e.g., spontaneous participation) or have hybrid formats. Moreira et al. (2023) studied patients on the AD continuum using a listening-based hybrid program combining live and recorded music tailored to participants’ preferences over several weeks. They found that memory improved in the experimental group compared with the control group, which only received standard care. The protocol was mainly receptive but included active elements intentionally designed to support a different cognitive goal in each session. These guided activities targeted specific cognitive functions such as episodic and autobiographical memory, orientation, and attention.

Comparison of active and receptive therapies

Few studies directly compare active and receptive music therapy, but available evidence generally favors active approaches. Gómez-Gallego et al. (2021) examined the effects of active and receptive music therapy in 90 patients with AD. Participants were randomly assigned to an active music therapy group, a receptive music therapy group, or a control group, and intervention sessions were held twice a week for three months. Active music therapy included rhythmic exercises, dancing, music quizzes, and music listening, whereas receptive music therapy focused on music listening combined with emotional or memory reflection guided by a facilitator. The authors reported cognitive improvement (MMSE) in 85.7% of patients in the active music therapy group, compared with 11.8% in the receptive therapy group and 6.3% in the control

group. Active music therapy also led to greater improvements in behavior and daily functioning than receptive therapy. Similarly, in the study by Särkämö et al. (2014), active music therapy (singing) showed larger improvements in certain memory functions than receptive music therapy, whereas receptive music therapy showed more reliable benefits for quality of life.

In addition to individual trials, a recent network meta-analysis by Ting et al. (2024) compared different music therapy formats for dementia and AD, including active music therapy, active music therapy with singing, rhythmic music therapy, music listening, and standalone singing (based on a single choir-based study). When cognitive outcomes were pooled across randomized controlled trials (RCTs), rhythmic music therapy, active music therapy with singing, and active music therapy without singing all showed statistically significant benefits for cognitive function compared with usual care. Among these approaches, rhythmic music therapy and active music therapy with singing showed the largest effects. In contrast, music listening and standalone singing were linked to smaller and potentially non-significant effects. Finally, a systematic review of RCTs by Bleibel et al. (2023) reported that patients who received active music therapy more often showed greater cognitive improvements than those who received receptive music therapy.

Taken together, these findings suggest that active music therapy approaches that include rhythmic or singing activities tend to produce greater cognitive improvements than receptive music therapy based mainly on music listening.

Music Selection: Individualized vs. Non-individualized Music

Another potentially significant factor in the effectiveness of music therapy is the selection of music. Personalized music selection may strengthen the benefits of music therapy in AD patients by engaging autobiographical memory and increasing emotional involvement (Gerdner, 2012; Sakamoto et al., 2013).

What is individualized music?

Individualized music refers to familiar, preferred, or personally meaningful songs from a person's past, usually identified through interviews with patients and caregivers. These songs often include music that was popular during the person's youth or connected to important life events. In contrast, non-individualized music consists of pieces chosen without direct personal relevance and is typically selected by the therapist based on general features such as genre, tempo, mood, and calming qualities. Many music therapy studies in AD use some degree of personalization, for example, by selecting music that matches the patient's general taste or music that was popular during their youth. This approach often aims to engage interest and activate remote memories, which tend to remain better preserved than recent memory in AD (Gerdner, 2012; Sakamoto et al., 2013).

Evidence for individualized vs. non-individualized music

Across studies, both individualized and non-individualized music interventions have been shown to improve symptoms of AD. Individualized music appears to increase engagement and to support both overall cognition and specific cognitive functions, such as autobiographical and episodic memory, when used regularly (Särkämö et al., 2014; Pérez-Ros et al., 2019; Arroyo-Anlló et al., 2013). Although one might expect non-individualized music to be less effective because it lacks personal meaning, studies show that non-individualized music interventions can still be beneficial when combined with active components such as rhythm, movement, or singing, or with social engagement such as group singing, and when repeated regularly (Cheung et al., 2018; Gómez-Gallego et al., 2021).

However, meta-analysis suggests that individualized music tends to produce larger cognitive effects regardless of the specific type or format of music therapy (Leggieri et al., 2019). When researchers directly compared individualized and non-individualized music within the same study, familiar music helped stabilize or improve self-consciousness and autobiographical memory in AD patients, whereas the unfamiliar music group showed a cognitive decline (MMSE) over the same period (Arroyo-Anlló et al., 2013). In experimental research focused specifically on autobiographical memory recall, AD patients recall more autobiographical memories when listening to personally chosen, meaningful music than when listening to researcher-chosen music or sitting in silence, while healthy controls did not show the same autobiographical benefit (El Haj et al., 2015). In addition, research suggests that music remains a strong cue for autobiographical memory compared to other types of cues. Baird et al. (2018) reported that the frequency of music-evoked autobiographical memories in AD patients was comparable to that of healthy controls, whereas photograph-evoked autobiographical memories were reduced in the AD group. Taken together, these findings suggest individualized music interventions may be more effective than non-individualized approaches in supporting cognitive function in AD.

Music Therapy vs. Other Non-Pharmacological Therapies

While music therapy shows promise for improving cognitive function in AD, it is also important to evaluate its effectiveness in comparison with other non-pharmacological interventions. Other therapies, including art therapy, physical activity, and meditation, are also commonly used to reduce symptoms of AD. However, there has been limited research that directly compares the effects of music therapy with these interventions on cognitive outcomes in AD. By examining how these approaches differ in their features or mechanisms, as well as their effects on cognitive functioning, we may better understand the unique benefits and strengths of music therapy in supporting cognition for individuals with AD.

Music therapy vs. painting

A study by Pongan et al. (2017) directly compared singing and painting interventions in mild AD patients over 12 weekly sessions. Patients in the singing group took part in group singing and the performance of well-known songs, whereas patients in the painting group

engaged in art discussions and painted together according to a predetermined theme. Both groups showed improvement in pain, working memory, inhibitory control, quality of life, and self-esteem. However, the singing intervention was more beneficial to episodic verbal memory, while the painting intervention led to greater reduction in anxiety and depression. Although it might be difficult to make a fully direct comparison between two very different activities, these findings suggest that music therapy may provide stronger cognitive benefits compared to visual art activities.

Music therapy vs. meditation

Innes et al. (2018) compared receptive music listening with Kirtan Kriya meditation in older adults with MCI or at risk for AD. The music-listening group listened to relaxing instrumental music from one of six classical composers selected by the participant. The meditation group practiced Kirtan Kriya, a meditation technique that involves repeating a mantra while coordinating fingers and visualizing mental imagery. Both groups showed improvements in memory function, cognitive performance, sleep, mood, and quality of life after 12 weekly sessions. However, the meditation group demonstrated greater and more consistent benefits, particularly in stress reduction, mood, and cognitive gains at the six-month follow-up. Biologically, the meditation group exhibited a significant increase in plasma amyloid- β ($A\beta_{40}$) and greater improvements in telomerase activity, suggesting potential amyloid- β clearance and anti-aging effects that may slow cognitive decline. The biomarker changes in the music-listening group were less pronounced compared to the meditation group.

However, these findings do not necessarily indicate that music therapy is less effective than meditation. Instead, the key difference likely lies in the level of engagement. In this study, Kirtan Kriya meditation involved more active engagement, requiring participants to actively produce chants or sounds, coordinate physical movements with their words, and visualize mental imagery. These elements likely stimulated multiple brain systems, including attention networks, language centers, motor coordination, and sensory systems. In contrast, the music intervention in this study involved passive listening only, which demands less mental effort and engagement. This pattern parallels observations from comparisons between active and receptive music therapy, in which active interventions that require higher levels of participation and engage more brain systems tend to produce greater cognitive benefits.

How Music Therapy Supports Cognitive Function in Alzheimer's Disease

To understand why music therapy can support cognitive function in AD patients, it is important to examine the underlying mechanisms through which music engages the brain. Music does not act on a single cognitive system. Instead, it activates multiple brain networks at the same time, which may be especially beneficial for AD patients whose brains often show significant damage in some regions while other networks remain relatively preserved.

Multimodal brain activation and preserved neural networks

Music therapy is unique because it engages multiple brain systems at the same time. Music simultaneously activates auditory, motor, attention, emotion, and reward systems across the brain (Jacobsen et al., 2015; Janata, 2009; Peck et al., 2016; Salimpoor et al., 2011). For example, listening to music activates the auditory cortex, while rhythms activate motor-related regions like the basal ganglia. Music can also stimulate brain areas involved in emotional processing, including the amygdala. In addition, research by Janata (2009) has shown that personally meaningful songs activate the autobiographical memory network in the medial prefrontal cortex, which plays a key role in integrating memory, emotion, and musical experience. Through this widespread, multimodal activation, music therapy may engage several cognitive systems at once in AD patients.

In the early stages of AD, the disease typically affects the medial temporal lobe, particularly the hippocampus and entorhinal cortex, which are critical for forming new episodic memories. As the condition progresses, neurodegeneration extends to association cortices in the temporal and parietal lobes, eventually impacting frontal regions responsible for higher-order executive functions. In contrast, many brain regions associated with music processing appear to remain relatively preserved, including the auditory cortex, motor areas, and limbic structures involved in rhythm, emotion, and reward (Baird & Samson, 2009; Jacobsen et al., 2015). Musical memories rely on brain regions that are often spared until later stages of the disease, which may help explain why music can continue to evoke strong emotional responses and autobiographical memories, or why songs from a long time ago are often remembered in AD patients. Because these music-related areas remain relatively intact, music may bypass damaged memory systems and access preserved neural pathways, making it a particularly effective stimulus for cognitive support in AD.

Autonomic nervous system and neurochemical modulation

Music has the potential to modulate the autonomic nervous system in ways that support cognitive function. Emotional feelings aroused by music tend to cause physiological changes due to automatic nervous system stimulation, leading to changes in heart rate, respiration, blood pressure, and overall arousal (Khalfa et al., 2002; Knight & Rickard, 2001). These physiological responses can influence the psychological state of patients, reducing anxiety and improving mood, which may further improve attention and memory performance (Nikkhah Bahrami et al., 2024; Peck et al., 2016).

Music therapy may also benefit AD patients through neurochemical mechanisms, particularly through engagement of the dopamine system (Chanda & Levitin, 2013; Peck et al., 2016). Dopamine is known for its involvement in reward processing, motivation, pleasure, and memory consolidation, which can aid in learning and memory retention (Wise, 2004). While more research on the neurochemistry of music is needed, diverse neurochemical systems beyond dopamine may contribute to an environment that supports cognitive engagement and mood regulation.

Limitations

Music therapy studies show several important limitations that restrict how confidently their findings on cognitive function in AD can be interpreted. These limitations relate to how interventions are designed, how outcomes are measured, and how participants are selected and compared across studies.

Duration and temporal effects

Most studies have implemented music therapy over relatively short periods, typically lasting 8–12 weeks, with sessions taking around 30–45 minutes. Such durations may be insufficient to capture or realize the full potential of music therapy. Studies conducted over longer timeframes may provide clearer information about whether music therapy improves cognitive symptoms in AD. It is also possible that meaningful cognitive benefits require prolonged exposure before they become detectable.

Additionally, evidence suggests that the cognitive benefits of music therapy may have temporal limits. Bruer et al. (2007) reported that AD patients showed cognitive improvement the day after active therapy sessions, but not the week after, indicating that the effects may fade quickly once the intervention ends (Bruer et al., 2007). This finding highlights the importance of implementing music therapy in a regular, frequent, and long-term manner to better support cognitive function in individuals with AD.

Measurement tools and timing of assessment

A viable cause for inconsistency among the results of the studies is the varying testing methods that different researchers employed. Cognitive outcomes were assessed using a wide range of testing tools, from global screening measures like the MMSE to more specific tasks targeting domain-specific functions such as memory and executive function (Bleibel et al., 2023). These tools differ in sensitivity and scope, which makes it difficult to directly compare results across studies. Some tools may be better at detecting subtle changes in cognition than others.

In addition, studies differed in how often and when cognitive assessments were conducted. For example, some studies administered assessments only at the baseline and immediately after the intervention, others tested repeatedly throughout the intervention, and some included follow-up testing weeks after the intervention ended. This variability can complicate comparisons and interpretation, as short follow-up periods may overestimate benefits by capturing only immediate or short-lived effects, while limited mid-point assessments make it difficult to determine when improvements emerge or whether they are maintained over time.

Participant selection and sample characteristics

Participant selection introduced further limitations. Many studies relied on small and heterogeneous samples, with substantial variation in age, disease severity, and comorbidities, which makes it harder to attribute observed cognitive changes specifically to music therapy

(Moreno-Morales et al., 2020; van der Steen et al., 2025). Several studies also included samples in which the majority of participants were women, leaving questions about whether the findings can be generalized equally to men and women with AD. In addition, because all participants already had some degree of cognitive impairment, difficulties with attention, comprehension, or engagement may have interfered with accurate assessment, further complicating the interpretation of results.

Control conditions

Finally, differences in control conditions and comparison groups limit the strength of conclusions about the cognitive effects of music therapy. Some studies lacked well-matched control groups, while others used a wide range of comparison conditions, including usual care, non-musical activities, or simple rest. These control conditions differ in their levels of social interaction and cognitive stimulation. As a result, it is difficult to determine whether observed cognitive improvements are specifically due to the music intervention itself or instead reflect more general effects of social interaction or other nonspecific factors associated with structured activities. Taken together, these methodological differences limit comparability across studies and weaken the overall evidence for the effects of music therapy on cognitive function in AD.

Future Directions

The findings of this review suggest several directions for future research on music therapy and cognitive function in AD. Further research is needed to clarify how music therapy should be delivered to achieve the greatest cognitive benefit. For example, studies could directly compare active music therapy with receptive listening-based approaches, as active therapies may engage multiple brain systems more strongly than receptive formats. Researchers should also examine whether individualized, autobiographically meaningful music leads to greater cognitive improvements than non-individualized or generic playlists by more powerfully activating memory and emotional networks.

To address current methodological limitations, future research should aim for longer intervention periods with appropriate session frequency and include follow-up assessments extending several months beyond the intervention to determine whether cognitive benefits are sustained over time. In addition, the use of standardized assessment protocols, larger and more diverse sample sizes, and rigorous randomized controlled designs would be essential for improving the reliability and generalizability of findings.

Finally, future studies should investigate the neurobiological mechanisms underlying the cognitive effects of music therapy using neuroimaging techniques such as fMRI or EEG, including measurements of brain activity before and after interventions. Understanding how music engages and potentially strengthens preserved brain networks for memory, attention, and emotion in Alzheimer's patients would provide stronger support for its clinical use.

Works Cited

- Alzheimer's Disease International. "World Alzheimer Report 2024." Alzheimer's Disease International (ADI), 2024, <https://www.alzint.org/u/World-Alzheimer-Report-2024.pdf>. Accessed 13 September 2025.
- American Music Therapy Association. "What is Music Therapy? | What is Music Therapy?" American Music Therapy Association, 2010, <https://www.musictherapy.org/about/musictherapy/>. Accessed 13 September 2025.
- Arroyo-Anlló, Eva M., Juan Poveda Díaz, and Roger Gil. "Familiar Music as an Enhancer of Self-Consciousness in Patients with Alzheimer's Disease." *BioMed Research International*, vol. 2013, no. 752965, 2013, <https://doi.org/10.1155/2013/752965>.
- Baird, Amee, and Séverine Samson. "Memory for music in Alzheimer's disease: unforgettable?." *Neuropsychology Review*, vol. 19, no. 1, 2009, pp. 85-101, <https://doi.org/10.1007/s11065-009-9085-2>.
- Baird, Amee, et al. "Characterization of music and photograph evoked autobiographical memories in people with Alzheimer's disease." *Journal of Alzheimer's Disease*, vol. 66, no. 2, 2018, pp. 693-706, <https://doi.org/10.3233/JAD-180627>.
- Bleibel, Malak, et al. "The effect of music therapy on cognitive functions in patients with Alzheimer's disease: a systematic review of randomized controlled trials." *Alzheimer's Research & Therapy*, vol. 15, no. 1, 2023, <https://doi.org/10.1186/s13195-023-01214-9>.
- Bruer, Robert, et al. "The Temporal Limits of Cognitive Change from Music Therapy in Elderly Persons with Dementia or Dementia-Like Cognitive Impairment: A Randomized Controlled Trial." *Journal of Music Therapy*, vol. 41, no. 4, 2007, pp. 308-328, <https://doi.org/10.1093/jmt/44.4.308>.
- Chanda, Mona Lisa, and Daniel J. Levitin. "The Neurochemistry of Music." *Trends in Cognitive Sciences*, vol. 17, no. 4, 2013, pp. 179-93, <https://doi.org/10.1016/j.tics.2013.02.007>.
- Cheung, Daphne Sze Ki, et al. "The effects of the music-with-movement intervention on the cognitive functions of people with moderate dementia: a randomized controlled trial." *Aging & Mental Health*, vol. 22, no. 3, 2018, pp. 306-315, <https://doi.org/10.1080/13607863.2016.1251571>.
- El Haj, Mohamad, et al. "Self-defining memories during exposure to music in Alzheimer's disease." *International Psychogeriatrics*, vol. 27, no. 10, 2015, pp. 1719-1730, <https://doi.org/10.1017/S1041610215000812>.
- Gerdner, Linda A. "Individualized music for dementia: Evolution and application of evidence-based protocol." *World Journal of Psychiatry*, vol. 2, no. 2, 2012, pp. 26-32, <https://doi.org/10.5498/wjp.v2.i2.26>.
- Gómez-Gallego, María, and Juan G. Gómez-García. "Music therapy and Alzheimer's disease: Cognitive, psychological, and behavioural effects." *Neurología*, vol. 32, no. 5, 2017, pp. 300-308, <https://doi.org/10.1016/j.nrl.2015.12.003>.
- Gómez-Gallego, María, et al. "Comparative Efficacy of Active Group Music Intervention versus Group Music Listening in Alzheimer's Disease." *International Journal of Environmental*

- Research and Public Health*, vol. 18, no. 15, 2021, <https://doi.org/10.3390/ijerph18158067>.
- Innes, Kim, et al. "Effects of Meditation and Music-Listening on Blood Biomarkers of Cellular Aging and Alzheimer's Disease in Adults with Subjective Cognitive Decline: An Exploratory Randomized Clinical Trial." *Journal of Alzheimer's Disease*, vol. 66, no. 3, 2018, pp. 947-970, <https://doi.org/10.3233/JAD-180164>.
- Jacobsen, Joˆrn-Henrik, et al. "Why musical memory can be preserved in advanced Alzheimer's disease." *Brain*, vol. 138, no. 8, 2015, <https://doi.org/10.1093/brain/awv135>.
- Janata, Petr. "The Neural Architecture of Music-Evoked Autobiographical Memories." *Cerebral Cortex*, vol. 19, no. 11, 2009, pp. 2579–94, <https://doi.org/10.1093/cercor/bhp008>.
- Khalfa, St phanie, et al. "Event-related skin conductance responses to musical emotions in humans." *Neuroscience Letters*, vol. 328, no. 2, 2002, pp. 145-149, [https://doi.org/10.1016/S0304-3940\(02\)00462-7](https://doi.org/10.1016/S0304-3940(02)00462-7).
- Knight, Wendy EJ, and Nikki S. Rickard. "Relaxing music prevents stress-induced increases in subjective anxiety, systolic blood pressure, and heart rate in healthy males and females." *Journal of Music Therapy*, vol. 38, no. 4, 2001, pp. 254-272, <https://doi.org/10.1093/jmt/38.4.254>.
- Leggieri, Melissa, et al. "Music intervention approaches for Alzheimer's disease: A review of the literature." *Frontiers in Neuroscience*, vol. 13, no. 132, 2019, <https://doi.org/10.3389/fnins.2019.00132>.
- Lyu, Jihui, et al. "The effects of music therapy on cognition, psychiatric symptoms, and activities of daily living in patients with Alzheimer's disease." *Journal of Alzheimer's Disease*, vol. 64, no. 4, 2018, pp. 1347-1358, <https://doi.org/10.3233/JAD-180183>.
- Moreira, Shirlene Vianna, et al. "Music Therapy Enhances Episodic Memory in Alzheimer's and Mixed Dementia: A Double-Blind Randomized Controlled Trial." *Healthcare*, vol. 11, no. 22, 2023, <https://doi.org/10.3390/healthcare11222912>.
- Moreno-Morales, Celia, et al. "Music therapy in the treatment of dementia: A systematic review and meta-analysis." *Frontiers in Medicine*, vol. 7, no. 160, 2020, <https://doi.org/10.3389/fmed.2020.00160>.
- National Institute on Aging. "Alzheimer's Disease Fact Sheet." National Institute on Aging, 5 April 2023, <https://www.nia.nih.gov/health/alzheimers-and-dementia/alzheimers-disease-fact-sheet#signs-and-symptoms-of-alzheimer-s-disease>. Accessed 13 September 2025.
- National Institute on Aging. "How Is Alzheimer's Disease Treated?" National Institute on Aging, 12 September 2023, <https://www.nia.nih.gov/health/alzheimers-treatment/how-alzheimers-disease-treated>. Accessed 13 September 2025.
- Nikkhah Bahrami, Sahar, et al. "Music Therapy for Alzheimer's Disease Management: A Narrative Review." *Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, vol. 60, no. 66, 2024, <https://doi.org/10.1186/s41983-024-00836-6>.

- Peck, Katlyn J., et al. "Music and memory in Alzheimer's disease and the potential underlying mechanisms." *Journal of Alzheimer's Disease*, vol. 51, no. 4, 2016, pp.949-959, <https://doi.org/10.3233/JAD-1509>.
- Pérez-Ros, Pilar, et al. "Preferred music listening intervention in nursing home residents with cognitive impairment: A randomized intervention study." *Journal of Alzheimer's Disease*, vol. 70, no. 2, 2019, pp. 433-442, <https://doi.org/10.3233/JAD-190361>.
- Pongan, Elodie, et al. "Can musical or painting interventions improve chronic pain, mood, quality of life, and cognition in patients with mild Alzheimer's disease?." *Journal of Alzheimer's Disease*, vol. 60, no. 2, 2017, <https://doi.org/10.3233/jad-170410>.
- Sakamoto, Mayumi, et al. "Comparing the effects of different individualized music interventions for elderly individuals with severe dementia." *International Psychogeriatrics*, vol. 25, no. 5, 2013, pp. 775-784, <https://doi.org/10.1017/s1041610212002256>.
- Salimpoor, Valorie N., et al. "Anatomically distinct dopamine release during anticipation and experience of peak emotion to music." *Nature Neuroscience*, vol. 14, no. 2, 2011, pp. 257-262, doi:10.1038/nn.2726.
- Särkämö, Teppo, et al. "Cognitive, emotional, and social benefits of regular musical activities in early dementia: randomized controlled study." *The Gerontologist*, vol. 54, no. 4, 2014, pp. 634-650, <https://doi.org/10.1093/geront/gnt100>.
- Ting, Berne, et al. "Multifaceted music therapy for depression in dementia: a network Meta-analysis of randomized controlled trials." *European Journal of Investigation in Health, Psychology and Education*, vol. 14, no. 2, 2024, pp. 351-367, <https://doi.org/10.3390/ejihpe14020024>.
- Tsoi, Kelvin KF, et al. "Receptive music therapy is more effective than interactive music therapy to relieve behavioral and psychological symptoms of dementia: a systematic review and meta-analysis." *Journal of the American Medical Directors Association*, vol. 19, no. 7, 2018, pp. 568-576, <https://doi.org/10.1016/j.jamda.2017.12.009>.
- Van der Steen, Jenny T., et al. "Music-based therapeutic interventions for people with dementia." *Cochrane Database of Systematic Reviews*, vol. 3, no. CD003477, 2025, <https://doi.org/10.1002/14651858.CD003477.pub5>.
- Wang, Zengmian, et al. "Music therapy improves cognitive function and behavior in patients with moderate Alzheimer's disease." *International Journal of Clinical and Experimental Medicine*, vol. 11, no. 5, 2017, pp. 4808-4814, <http://www.ijcem.com/files/ijcem0048744.pdf>.
- Wise, Roy A. "Dopamine, learning and motivation." *Nature Reviews Neuroscience*, vol. 5, no. 6, 2004, pp. 483-494, <https://doi.org/10.1038/nrn1406>.

A Nation Ignited: The Origins of Orbit and the Public's Ascent

By Sienna Tolani

INTRO

On October 10th 1960, readers of *Missiles and Rockets* magazine, accustomed to leisurely articles on a topic of interest, were met with a dire warning from United States Senator John F. Kennedy, then the Democratic candidate for president. "Control of space will be decided in the next decade," Kennedy proclaimed, "If the Soviets control space, they can control Earth, as in past centuries the nation that controlled the seas has dominated the continents." "We cannot run second in this vital race," he continued. "To ensure peace and freedom, we must be first."¹⁴³

Kennedy's alarming declaration, splashed across the flimsy pages of a popular magazine, begged the American public to recognize the urgent need to reorganize the American defense establishment and prioritize the space program. By 1960, the US had been losing the Space Race against the Soviet Union for three years, and, despite continued efforts to further space development, there were no signs of an imminent American victory. If Kennedy's words were to be believed, in 1960 Soviet success in space had left it effectively in control of the planet. The only way to escape Soviet domination, therefore, was for the United States to catch up to Soviet achievements in space, and then move further by conquering the moon.

The "Space Race" was a twelve year period of intense competition between the United States and the Soviet Union for supremacy in space exploration. Ultimately, it resulted in unprecedented advances in aerospace technology in both countries. It began in 1957 when the Soviets launched Sputnik-I, the world's first artificial satellite, and it effectively ended when the U.S. emerged victorious with its 1969 moon landing. Competition in space was part of a larger Cold War struggle being waged between the two superpowers for global dominance. In the words of historian Karsten Werth, the Space Race served as "a surrogate for war."¹⁴⁴ Just as with the other proxy wars that the two nations pursued during the Cold War, victories or losses in the Space Race had propaganda potential that could enhance or reduce the prestige of each side internationally.

Ironically, it was neither American nor Soviet technology that set the stage for this competition. It was Germany's formidable rocket and missile technology, unleashed in the final phase of WWII, that kindled the first flames. In 1944 and 1945, as the Third Reich collapsed, invading Soviet and American forces captured German resources and used them to accelerate their own ambitions in rocketry, culminating in the Soviet Union's victorious launch of Sputnik. The launch of Sputnik had profound effects in the United States, in both the short and longer term. The initial response of the people and government was widespread panic followed by a dramatic series of federal policy overhauls that reshaped the American education system and redirected federal spending. In the longer term, Sputnik and the Space Race that followed

¹⁴³ "Missiles and Rockets: Sept. 7 - Dec. 1960," Internet Archive, last modified January 8, 2024, accessed May 23, 2025, <https://archive.org/details/missilesrockets7196unse/page/n341/mode/2up>.

¹⁴⁴ Karsten Werth, "A Surrogate for War—The U.S. Space Program in the 1960s," *Amerikastudien / American Studies* 49, no. 4 (2004): 1, JSTOR.

inspired a new area of popular culture focused on space exploration. In 1961, in an effort to restore American confidence and divert the public's attention from recent Cold War failures, President John F. Kennedy championed the Apollo Project, attempting to rally Americans behind an extraordinary lunar mission. While Kennedy succeeded in igniting the patriotism of the American public at first, the later turbulence of the 1960s, ranging from the Vietnam War to the War on Poverty, eventually eroded the public's support of the Space Race and prompted a mixed response to Apollo.

There is little consensus among historians regarding the U.S. motive for entering the Space Race. This debate is evidenced by the conflicting views of historians Amy Paige Snyder and Susan Landrum Mangus. Mangus contends that the United States developed space technology and engaged in a space race with the USSR due to "Americans' fascination with the frontier". In her argument, Mangus explains that the concept of the frontier encompassed more than the typical symbols of cowboys moving westwards, but spanned back to the initial English settlement in America. Confronting the frontier was "a feat performed by 'courageous' men and women". This narrative aligned with the United States' desires to outperform the Soviet Union in space, asserting its superiority and "courageousness". Mangus notes that NASA leadership employed frontier terminology when presenting the Space Race to the public; astronauts were painted as "courageous" pioneers, like the frontiersmen who migrated west during the Age of Manifest Destiny or the first Europeans to travel across oceans to reach the "New World". By this thesis, the Space Race was another "challenging adventure" for Americans to embark on, similar to those they had succeeded in, in the past. Finally, Mangus states that Americans recognized the significance of the frontier to national interests and believed that the Space Race could engender the same effects of "rejuvenating the nation and safeguarding democracy for future generations". She cites President Kennedy's use of the word "frontier" when describing America's goals in science and technology in the Space Race.

Explaining America's decision to venture into space from a starkly different perspective, the historian Amy Paige Snyder argues that, while the Cold War contributed to the Space Race, "earlier efforts in science provided the critical foundation that enabled the American space program's initiation". To this effect, Snyder recounts various historical discoveries throughout the study of astronomy and rocketry, asserting that these breakthroughs, driven solely by scientific inquiry, are what set the pretext for the Space Race. Some developments that Snyder notes are Tsiolkovsky's first theoretical foundations, Hermann Oberth's publication of "*Die Rakete zu den Planetenräumen* (The Rocket into Interplanetary Space), and Robert Goddard's homemade rockets. Snyder explains that these early advances in rocketry are what inspired the creation of American satellites in the first place.¹⁴⁵ Although it is true that American rocketry developed from the foundations established by scientists in the earlier half of the 20th century, Snyder fails to stress the most important of these: the Nazis creation and development of the V-2, and America's eventual theft of Germany's war tech.

¹⁴⁵ "Space Race: What Prompted the United States to Venture into Space?," in *American Social and Political Movements, 1945-2000: Pursuit of Liberty*, ed. Robert J. Allison, vol. 2, *History in Dispute* (Detroit, MI: St. James Press, 2000), digital file.

SECTION I: THE ROOTS OF ROCKETRY

Prior to World War II, individual Americans indulged in rocket research at their own whims. Their projects occurred at a slow pace and were driven by their individual interests. WWII, however, marked a turning point in rocket research. During the war, Germany unveiled its formidable and lethal V-2 missile. Upon realizing the potential of rocket and missile power, the United States and Soviet Union exploited their victory in WWII, hunting the remains of German missile and rocket research in order to bolster their own ambitions in rocketry. Their adoption of German research resulted in the first stages of the Space Race as both the US and USSR sought the means to harmlessly project and experiment with the power they contrived using German resources.

Throughout the 1930's, the memory of the financial and human cost of World War I had left both the American public and government officials adamant that the nation should avoid involvement in another war abroad. Americans felt that they did not need to spend on their military or concern themselves with politics in Europe because of the protection that their geographical position rewarded them.¹⁴⁶ With an ocean between the United States and Europe, and the emergence of the Great Depression, Americans concluded that government funding should be focused on aiding the nation's "actual needs", like the elimination of poverty. Thus, throughout the interwar period, the United States pursued a foreign policy centered around isolationism and disarmament. The American government deemed measures that would prepare the nation for war or increase its arm power unnecessary and threats to peace, leading to little funding being allocated for missile and rocket development.¹⁴⁷

In 1928, fifteen nations (eventually over seventy, once it was established) signed the Kellogg-Briand Pact which prohibited war "as an instrument of national policy". The pact was realized after French foreign minister Aristide Briand requested U.S. President Calvin Coolidge to agree to a treaty that would outlaw war between the two powers. When Coolidge refused to reply, the American public began rallying behind the idea of outlawing war. Secretary of State Frank Kellogg agreed that the idea was plausible if several nations signed the treaty.¹⁴⁸ Although the intentions of the pact were overly idealistic and eventually ignored, it exemplified the American desires for isolationism. After assuming office in 1933, President Franklin D. Roosevelt's administration furthered the antiwar agenda. In May 1933, the Disarmament Conference convened by the League of Nations occurred in Geneva. Addressing the conference, the U.S.'s chief delegate, Norman H. Davis, shared America's intentions to avoid warfare and to disarm "through successive stages down to the basis of a domestic police force". He articulated America's decisions to not invest in or construct new weaponry, "...weapons of an aggressive

¹⁴⁶ David E. Hudson, "DISARMAMENT URGED as AN ECONOMY MOVE: THE UNITED STATES IS VIEWED as HAVING SPLENDID OPPORTUNITY to save SAFELY," *New York Times* (1923-) (New York, N.Y.), January 22, 1933, ProQuest Historical Newspapers: The New York Times (100599886).

¹⁴⁷ Special to THE NEW YORK TIMES, "Peace and Arms Pledges of the United States Given to the World Disarmament Parley by Davis," *New York Times* (1923-) (New York, N.Y.), May 23, 1933, ProQuest Historical Newspapers: The New York Times (100649182).

¹⁴⁸ Sonia Benson, Daniel E. Brannen, Jr., and Rebecca Valentine, "Kellogg-Briand Pact," in *UXL Encyclopedia of U.S. History* (Detroit, MI: UXL, 2009), 5 Gale in Context: U.S. History.

character...are the more costly to construct and...are those most likely to lead to a breach of the peace.”¹⁴⁹ Levels of U.S. defense spending throughout the 1930s reveals that this was not just rhetoric. Although gradual increases occurred on the eve of World War II, most of the decade saw annual expenditure ranging between 13,096 to 23,129 millions of dollars. By contrast, defense spending at the peak of WWII ranged from \$1,052,686 to \$1,246,107 million.¹⁵⁰

The U.S. commitment to isolationism and disarmament was also asserted in the passage of three Neutrality Acts. The Neutrality Act of 1935, signed into law on August 31st, banned all shipments of arms and ammunition to aggressor nations, thus preventing American involvement in conflicts like Italy’s invasion of Ethiopia, which occurred later that year. By not providing arms to foreign powers, American manufacturers were encouraged to halt weapon production, and nations that could not generate their own arms were forced to avoid warfare. A majority of Americans supported the law. A November 1935 Washington Post poll revealed that 47% of those polled wanted to “prohibit all trade with nations at war” and a like minded 37% felt the government should “prohibit trade in war materials only”. In 1936, the second Neutrality Act was passed, extending the first act by fourteen months and adding a provision that outlawed private lending to belligerent nations. The Neutrality Act of 1937 made the act of 1936 permanent, but added a caveat to ensure American profits; for a period of two years, American companies could sell weapons to nations at war, provided they paid in cash and transported the goods on vessels that were unaffiliated with the U.S..¹⁵¹

Throughout this period, American research in rocket science remained in its infancy, lacking both public interest and funding. There were a few exceptions, however. Most notably, Robert Goddard, developed the first liquid fuel rockets in 1926. With the aid of a Smithsonian grant, Goddard was able to develop a homemade liquid fueled rocket and launch it 56m high from his backyard. Afterwards, Goddard moved to New Mexico to escape from journalists, competing rocketeers, and neighbors. There he made even more progress, performing about one-hundred static tests, launching forty-eight large experimental rockets and receiving eighty-three patents for his work. An absence of greater funding and a seeming unwillingness to collaborate with other scientists prevented him from achieving many of his more ambitious goals, but his final rockets were still able to produce a thrust of up to eight hundred and twenty-five pounds.¹⁵²

It was only later, in 1939, that American rocket pioneers began receiving substantial funding from the federal government. General “Hap” Arnold, the commander of the Army Air Corps, requested R&D sponsorship for the development of small rockets that would assist in the takeoff of heavily loaded planes. His request secured the first \$1,000 million for Caltech’s Guggenheim Aeronautical Laboratory. Similarly, in 1940, the establishment of the Office for

¹⁴⁹ Special to THE NEW YORK TIMES, "Peace and Arms,".

¹⁵⁰ Robert Johnston, "U.S. Expenditures for Defense and Education, 1910-2021," table, Johnston's Archive, January 29, 2023, accessed May 22, 2025, <https://www.johnstonsarchive.net/policy/edgraph.html>.

¹⁵¹ David G. Delaney, "Neutrality Acts," in *Major Acts of Congress*, ed. Brian K. Landsberg (New York, NY: Macmillan Reference USA, 2004), 3:, Gale in Context: U.S. History.

¹⁵² Walter A. McDougall, *The Heavens and the Earth: A Political History of the Space Age* (New York: Basic Book, 1985), 77.

Scientific Research and Development (OSRD) under the leadership of former MIT Vice President Vannevar Bush resulted in federal funding for university research. With federal funding, the University of California and MIT founded Radiation Laboratories, Harvard founded a center for radar research, and the University of Chicago founded a nuclear Metallurgical Laboratory.¹⁵³

The American withdrawal from foreign affairs and armaments production, including rocket research, stood in contrast to the path followed by Germany and the Soviet Union during this period. The early thirties saw the formation of two rocket societies in the USSR: Group for the Study of Reactive Motion (GIRD) in 1931 and the ten-manned Gas Dynamics Laboratory (GDL) in 1928. GIRD primarily focused on rocketry as a means of spaceflight whereas the GDL explored its potential military applications.¹⁵⁴ In late 1932, the Red Army's chief of staff, Marshal Mikhail Tukachevskii took notice of the GDL's work and proposed combining it with the GIRD to create a single, nationally sponsored R&D institute. After a series of negotiations, the two groups were officially united on September 21st, 1933 by an order signed by Tukachevskii.¹⁵⁵ Nevertheless, arguments between the previously distinct groups resulted in an extremely slow-paced Soviet rocketry program. First, a debate surrounding whether to use solid or liquid propellants in Soviet rockets quickly unfolded. While the GDL had been utilizing solid propellants for over a decade, the GIRD favored liquid propellants. More conflict arose regarding the ideal choice of liquid propellant.¹⁵⁶ GIRD veterans opted to use liquid-oxygen engines as suggested by early rocket-visionary Tsiolkovskii, but the GDL researchers believed it best to utilize storable fuels like nitric acid.¹⁵⁷ This disagreement endured for years as members from both the GIRD and GDL factions wrote letters to government officials and the Communist party, attempting to discredit the other side's argument in favor of their preferred fuel. By the late 1930s, the RNII also encountered severe cuts in funding. In 1936, the Narkomtiazhprom, the ministry that administered RNII operations, only provided the RNII with 3.42 million rubles out of the 1.6 billion rubles it spent on defense.¹⁵⁸

A third blow was dealt to the RNII in 1937: Stalin's Purge. The Great Purge was a period of intense political repression in the Soviet Union. Anybody suspected of opposing Stalin's regime was immediately arrested and often killed. On May 26th, 1937, Marshal Tukachevskii was arrested alongside seven other senior officers. One of these officers included Robert Eidemann, Tukachevskii's aide who had sponsored GIRD's early work. All eight men arrested were subsequently shot on June 12th, 1937.¹⁵⁹ Occurring concurrently to the RNII's internal disagreements, members of the GIRD and GDL began employing the political turmoil to further their interests in rocket production. Ivan Kleimenov, a former GDL member and the director of

¹⁵³ McDougall, *The Heavens*, 79.

¹⁵⁴ Asif Siddiqi, "The Rockets' Red Glare: Technology, Conflict, and Terror in the Soviet Union," *Technology and Culture* 44, no. 3 (2003): 6-7, JSTOR.

¹⁵⁵ Siddiqi, "The Rockets," 8-9.

¹⁵⁶ Siddiqi, "The Rockets," 11.

¹⁵⁷ Siddiqi, "The Rockets," 12.

¹⁵⁸ Siddiqi, "The Rockets," 16.

¹⁵⁹ Siddiqi, "The Rockets," 21.

the RNII, was arrested and killed in January 1938 after Leonid Korneev, a pro-liquid-oxygen engineer who Kleimenov had fired, wrote to the Defense Commissioner, accusing Kleimenov of anti-Soviet activities. The executions of rocket engineers and administrators continued well into 1938, taking a toll on Soviet rocketry production prior to WWII.¹⁶⁰ Therefore, while the Soviet Union invested significantly more than the United States in rocket development throughout the 1930s, fatal internal disagreements stymied Soviet rocket efforts, resulting in the Soviets' substantial lag behind German rocketry.

German interest in rocket science took root just before the ascension of Adolf Hitler in 1933. Once he was in office, Hitler initiated a campaign to rearm Germany's military, overturning the peace settlement imposed by the Versailles Treaty and expanding the German rocket program. Hitler viewed Eastern Europe as an opportunity to build the German empire by land. In an effort to accomplish this goal, he prepared for future military conflicts by subtly rearming Germany's military throughout the 1930s. In 1933, he removed Germany from the League of Nations, therefore withdrawing it from the international consensus for disarmament. Later, in 1935, Hitler established the Luftwaffe (the German air force) and increased the size of Germany's army to surpass the 100,000 men limit that the Treaty of Versailles had set. Finally, in 1936, Hitler formally abrogated the Versailles Treaty and announced rearmament with a program known as the "Four Year Plan." The Four Year Program focused on developing Germany's heavy industry and arms.¹⁶¹

German rocket research began during the Weimar Republic. Though the German military had been made impotent by the Treaty of Versailles, Germans military leaders still sought a weapon system that was novel, lethal, and not off limits to them under the Treaty. In this light, rockets held the curious potential to serve as Germany's new weapon carriers.¹⁶² The Reich Board of Ordnance began its rocket research in 1930, although obtaining funding for its research and the facilities it required for experimental tests had been difficult. In fact, according to Walter Dornberger, a long-range artilleryman who coheaded the eventual creation of the V-2, research had only continued due to the support of General Karl Becker, chief of the development department of the Army Board of Ordnance.¹⁶³ Becker had to scrape together money from other divisions in order to fund Dornberger's research. Nonetheless, in 1936, as Germany commenced the Four Year Program, Becker explained to Dornberger that research could more easily continue if it proved of "military value". Thus, Dornberger and his team--Wernher von Braun, Walter Riedel, Heinrich Grunow--set out to construct "big operational rockets".¹⁶⁴

In May 1937, presumably with federal financial assistance, these four scientists, as well as an additional 300 men, moved into Peenemunde. Located on a large island in the Baltic,

¹⁶⁰ Siddiqi, "The Rockets," 22-23.

¹⁶¹ "Foreign Policy and the Road to War," *The Holocaust Explained*, accessed May 22, 2025, <https://www.theholocaustexplained.org/life-in-nazi-occupied-europe/foreign-policy-and-the-road-to-war/build-up-to-war/#:~:text=Conscription%20and%20Rearmament%201935,the%20Luftwaffe%20had%208000%20planes>.

¹⁶² Walter R. Dornberger, "The German V-2," *Technology and Culture* 4, no. 4 (1963): 394, <https://doi.org/10.2307/3101376>.

¹⁶³ Dornberger, "The German," 397.

¹⁶⁴ Dornberger, "The German," 397.

Peenemunde was a “self-sufficient research and development center for rockets”.¹⁶⁵ Once war broke out in 1939, Hitler gave minimal priority to operations at Peenemunde, harboring little faith that it would be able to develop any missiles capable of carrying bombs into London.¹⁶⁶ He was wrong. In 1943, years of collaboration by German industry and universities resulted in the development of the V2, the world’s first long-range ballistic missile.¹⁶⁷ In July 1943, Hitler was convinced of its power, but a German loss was unavoidable and could only be delayed. In this last stretch of the war, Hitler invested the diminishing Nazi funds into V-2 development. V-2 production was relocated to Mittelwerk Factory in the Mittelbau-Dora concentration camp where thousands of enslaved prisoners were put to work on its assembly. By 1945, an estimated 900 V-2s were being produced in Mittelwerk each month. More than 10,000 prisoners are believed to have died during the work. As historian Walter McDougall described the labor conditions of Mittelwerk, “If too weak to work, slaves were left to expire--150 per day--human sacrifices on the altar of machines and power.”¹⁶⁸

The first V-2 of WWII was launched on September 6th, 1944, against an Allied military camp outside of Paris. Two days later, several additional rockets were fired into London and parts of southern England.¹⁶⁹ By 1945, the Germans were employing the V-2 at a fatal rate, firing on England nearly every month. On April 27th, 1945, the Times of London published an article detailing the horrific attacks on England that had occurred since September 1944. In a span of seven months, a total of 1,050 rockets were estimated to have reached England, killing 2,754 people, and “seriously injuring” upwards of 6,500. Forty-five churches and chapels had been destroyed, 35 hospitals were hit, and schools and libraries were also turned to rubble.¹⁷⁰ England’s tragic fate showcased the unbelievable power of the V-2 on the world stage. The Allies, including both the United States and USSR, were made privy to the V-2’s potential.

As an imminent Allied victory became clear, Werner von Braun and several other prominent Nazi scientists, hoping to avoid arrest, offered their surrender to American officials in Nordhausen, Germany. Following their concession, Americans uncovered the Mittelwerk facility on April 11th, 1945, and discreetly scrambled to take the wholes and pieces of V-2 rockets.¹⁷¹ In July 1945, leaders of the “Big Three” (United States, United Kingdom, and Soviet Union) convened a conference in Potsdam. There, they agreed to share in the capture of German scientific facilities as a form of “intellectual reparations.” While the agreement gave the appearance of inter-allied cooperation on German weapons research, the reality was that both the United States and the Soviet Union had already confiscated the materials for themselves. The Soviets were “carting off” entire German laboratories and the U.S. had already shipped rockets

¹⁶⁵ Dornberger, "The German," 400.

¹⁶⁶ Dornberger, "The German," 397.

¹⁶⁷ Dornberger, "The German," 397; The Editors of Encyclopedia Britannica, "V-2 Rocket," Britannica, last modified May 2, 2025, accessed May 22, 2025, <https://www.britannica.com/technology/V-2-rocket>.

¹⁶⁸ McDougall, *The Heavens*, 42-44.

¹⁶⁹ Dornberger, "The German," 406.

¹⁷⁰ "London Ordeal of V2s," *The Times*, April 27, 1945, The Times Digital Archive.

¹⁷¹ McDougall, *The Heavens*, 44.

across the Atlantic.¹⁷² As both the U.S. and USSR set out to obtain the remains of German rocket research and the genius of German scientists, they clearly exhibited a fascination with and a desire for the rocket power that Germany had demonstrated with their V-2s.

In the fall of 1945, the United States launched Operation Paperclip, a secret program that recruited more than a hundred German scientists, including Von Braun, to relocate to the United States to help develop American missiles and rocketry.¹⁷³ As these Germans began to lead American rocket programs, in 1946 the think tank RAND Corporation started a campaign to convince the American government to undertake development of an artificial satellite, using its newfound German expertise. After this report failed to gain traction, RAND published a second report in 1950, once again attempting to push the United States military towards building a satellite. In the 1950 report, RAND more thoroughly detailed the positive military and psychological effects of a satellite. Primarily, RAND arguing that it could be utilized to observe hard-to-access regions around the world and to spy on enemies. Eventually, the Truman administration agreed that the United States should direct its efforts toward satellite development and began developing a public rationale in an effort to avoid foreign accusations of spying.

Providing a solution, in 1952, the International Council of Scientific Unions proposed the creation of the International Geophysical Year (IGY). The IGY would join together sixty-seven nations between 1957 and 1958, a two year period of maximum solar activity, in an effort to study earth science. In October 1954, the United States's plan for developing and launching an orbiting artificial satellite during the IGY was approved by the IGY's Special Committee. Later, in 1955, the Soviets' commission was approved as well. In May 1955, the National Security Council (NSC) submitted a top-secret report to President Eisenhower. In it, the council advised Eisenhower to support federal financing for America's mission to send a satellite into space due to the project's close technological association with intercontinental ballistic missiles. The NSC believed that if the United States succeeded in launching a satellite first, it would act as a symbol of American military prowess and discourage foreign nations from succumbing to the influences of communism in the greater Cold War, "...such a demonstration of advanced technology and its unmistakable relationship to intercontinental ballistic missile technology might have important repercussions on the political determination of free world countries to resist communist threats."¹⁷⁴ Despite this recognition, the US failed to beat the Soviets in satellite development.

SECTION II: THE AFTERMATH OF SPUTNIK-1

The Soviet launch of Sputnik on October 4th, 1957, provoked widespread panic among both the public and government officials. To many, the event served as evidence that the Soviets possessed the capability to surpass the United States in technological and military prestige. It also dealt the Soviets a propaganda victory at a time when the U.S. sought to portray itself to the developing world as an unequivocal political and technological leader. The Eisenhower

¹⁷² McDougall, *The Heavens*, 45.

¹⁷³ McDougall, *The Heavens*, 45.

¹⁷⁴ Mark Lawrence, "The Space Race and American Foreign Relations," in *The Oxford Encyclopedia of the History of American Foreign Relations* (Oxford University Press, 2021), Oxford Reference Online.

administration's reaction was swift and dramatic; within a year of Sputnik's launch, the government initiated a series of effective policies in the realms of space research and public education. The public's response was no less pronounced. Over the following decade, outer space captivated American popular culture, triggering a demand for space-related products and media, and the creation of youth rocketry clubs.

Despite the U.S. government's alarm, Sputnik-1 was a relatively simple device. Determined to claim the title of the first nation to place a satellite into orbit, the Soviets had tirelessly worked to design an intricate, scientific satellite. However, when development of this sophisticated device fell behind schedule, Chief Space Engineer Sergei Korolev opted for a less complex satellite that could be ready sooner. Measuring just two-feet in diameter, Sputnik-1 was an aluminum alloy sphere with four external antennas that emitted brief radio pulses that were audible to even the most amateur radio operators as iconic "beep-beeps." Beyond the radio, the satellite's reflective surface and large eighty-five-foot launch vehicle rendered it visible to eager spectators across the world. Sputnik was launched into orbit nearly 600 miles above the Earth, and reentered Earth's atmosphere on January 4th, 1958.¹⁷⁵ Although its orbit was brief, Sputnik-1 remained the fixation of global audiences for months following its launch, projecting an impressive image of Soviet ballistic missile abilities and casting a shadow on the United States leadership in technology and innovation. The international press incited both global and domestic doubts of the U.S., reporting on the Soviets' success and the power dynamics between the two superpowers. The Soviets' subsequent launch of the larger Sputnik-2, one month after the first Sputnik, only furthered these concerns. Sputnik-2 not only weighed six times as much as Sputnik-1, it also successfully carried the first living creature, a dog named Laika, into space.¹⁷⁶

On October 21st, 1957, *TIME* Magazine published an article titled "The Race to Come" in which it detailed a snapshot of the nation's shaken psyche. *TIME* writers described that, "The U.S. takes deep pride in its technical skills and technological prowess, in its ability to get things done--first...Now, there [is] a sudden, sharp national disappointment...[that will] linger until the U.S. no longer [stands] second best in the conquest of space."¹⁷⁷ Americans felt a collective wounding of national pride and a desire to score the next victory in the Space Race. This urgency was amplified by fears that the Russians would weaponize their rocket technology to attack the United States. Prominent scientists, possibly overdramatizing the situation in order to receive more public and financial support, animated these fears in press interviews. In a *Daily Defender* article dated November 25th, 1957, Edward Teller, known as the father of the H-bomb, warned that the United States could potentially remain behind the USSR for as much as ten years.¹⁷⁸ Similarly, in a *New York Times* interview published December 15th, 1957, renowned German rocket scientist Wernher Von Braun testified that the USSR was indeed capable of striking D.C.

¹⁷⁵ John M. Logsdon, "Sputnik," in *America in the World, 1776 to the Present: A Supplement to the Dictionary of American History*, ed. Edward J. Blum (Farmington Hills, MI: Charles Scribner's Sons, 2016), 2: Gale eBooks.

¹⁷⁶ Lawrence, "The Space,".

¹⁷⁷ "The Race to Come," *TIME Magazine* 70, no. 17 (1957): Time Magazine Archive.

¹⁷⁸ "Says U. S. Lags in Space Race: Says U. S. behind in Space Race," *Daily Defender (Daily Edition) (1956-1960)* (Chicago, Ill.), November 25, 1957, ProQuest Historical Newspapers: Black Newspaper Collection (493595615).

with an H-bomb, using its rockets. When asked by Senator Lyndon Johnson if the attack was plausible, Von Braun bluntly replied “Yes.”¹⁷⁹ These sensational falsehoods only intensified after Sputnik-2. A Gallup poll from mid-January 1958, weeks after Sputnik-2’s launch, revealed that 67% of the Americans surveyed believed that Russia was “ahead in the Cold War” due to Sputnik, and 35% of Americans believed “that the Russians could wipe out most cities in the U.S. in a matter of a few hours with their new rockets and missiles.” Americans feared a reality that was far from the truth. In fact, the Soviets were years away from producing long-range missiles of such strength and range.¹⁸⁰

One result of Sputnik’s launch was scrutiny of the American educational system. Many critics argued that the Russians had beaten the U.S. to space because of its superior math and science instruction in schools.¹⁸¹ These concerns were voiced by a diverse range of Americans, from average teenagers to government officials. Edward Teller emphasized that, in particular, young American girls needed to shift their aspirations to more closely resemble the scientific pursuits of their Russian counterparts.¹⁸² In a *Washington Post* article published after Sputnik-2’s launch, a group of young rocket enthusiasts aired their similar opinions on the matter. Teens Martin Fink, Beverly Brewster, Eddie Ganz, and Sharon Reagin, who excelled in math and science in their respective high schools, agreed that more high school students needed to “get in on the ground floor [and] learn math” to begin with. They lamented that many of their peers were “scared away” by the mere word “science”. In line with these comments, Dr. Leo Schubert of the Junior Academy of Sciences declared that “high school counselors...ruined a whole generation of students” by promoting that they could not learn math.¹⁸³ Sentiments that the Soviet school system instilled greater discipline and rigor in math and science continued to gain traction amongst the public and legislators. Individual schools took action, passing resolutions dedicating their institutions to academic excellence and scientific exploration by students. Even morning radio hosts admonished their teenage audiences, preaching that they needed to excel in their academics in order to prevent the nation from falling further behind the Russians.¹⁸⁴

Due to public pressure and his own rising concerns, President Eisenhower took steps to revamp America’s math and science education. Eisenhower viewed education as both a national security issue and a means to regain international prestige. He called for educational reform that would favor investments in math, science, and technological education, believing education to be a tool to mobilize American strength. In his eyes, with federal aid to education, the United States

¹⁷⁹ By JOHN D. MORRIS Special to The New York Times, "Bigger Rocket an Urgent Need, Von Braun and Medaris Warn: Missile Experts Declare U.S. Will Lose Race to Russians Unless It Acts Fast STRONGER ROCKET HELD VITAL to U.S. Von Braun Makes a Hit Keller's Role Praised," *New York Times* (1923-) (New York, N.Y.), December 15, 1957, ProQuest Historical Newspapers: The New York Times (114065439).

¹⁸⁰ Rodger A. Payne, "Public Opinion and Foreign Threats: Eisenhower's Response to Sputnik," *Armed Forces and Society* 21, no. 1 (1994): Page 93, JSTOR.

¹⁸¹ Lawrence, "The Space."

¹⁸² "Says U. S. Lags."

¹⁸³ By Winzola McLendon, "Teen Scientists Have Answers: They're Space Age 'Pilots,'" *The Washington Post and Times Herald* (1954-1959) (Washington, D.C.), November 17, 1957, ProQuest Historical Newspapers: The Washington Post (148839053).

¹⁸⁴ Homer H. Hickam, Jr, *Rocket Boys* (New York: Dell Publishing, 1998), Page 25.

would generate more accomplished scientists and appear stronger against the Soviet opposition.¹⁸⁵ In this spirit, Congress passed the National Defense Education Act in 1958. The act created a low-interest loan program worth \$295M for college students and poured \$300M into public schools to enhance science and math programs.¹⁸⁶ 24,831 students accepted federal loans in 1959 and a massive 247,000 students took advantage of their loans in 1964. These funds invigorated learning at all levels. In the first six years following the act's passage, federal funding improved or equipped approximately 3,400 elementary and 7,700 secondary school classrooms or labs.¹⁸⁷ Furthermore, American public schools deployed unconventional tools to improve their math curriculums. Originally tested in St. Louis high schools before being supplied nation-wide, the U.S. Office of Education imported Soviet eighth-grade level texts for secondary math classes. These math books, stripped of visual embellishments and focused solely on content, were well-liked by the American students. One American boy explained that, "U.S. books have lots of colored pictures. [The Soviet] one has only small, simple figures, but it gets right to the point."¹⁸⁸

The National Defense Education Act emerged within the context of other federal interventions in American life. In 1944, Congress passed the GI Bill, lending financial assistance to returning veterans seeking first-home mortgages. The same law provided veterans aid for higher education throughout the succeeding decade. Millions of families were now able to purchase their first homes and hundreds of thousands became the first in their family to attend college.¹⁸⁹ Given the government's expanding role in American lives, as evidenced by the GI Bill, its growing investment in R&D seemed natural. By the late 1950s, a majority of R&D dollars spent on basic research on rocketry came from the federal government. As early as 1953, the government was allocating \$153M to research while the universities that used the funds were contributing a fractional \$10M. By 1960, the federal government's spending on R&D had tripled to exceed 65% of the \$19.6B total.¹⁹⁰ The federal government spending on education was not only aimed at bolstering national security, however. It was also meant to drive the social and economic advancement of the American people. In this way, the Education Act epitomized the federal expansion typical of the period. Traditionally, education policy and financing had fallen strictly under the domain of state and local governments. The Education Act was the first federal attempt to engage society in a national effort to improve education.

The National Defense Education Act was only one component of Eisenhower's broader Sputnik strategy. Although Sputnik was a surprise to all Americans, Eisenhower admitted that he

¹⁸⁵ Sean Kay, "America's Sputnik Moments.," *Survival* (0039-6338) 55, no. 2 (2013): Page 3, <https://doi.org/10.1080/00396338.2013.784470>.

¹⁸⁶ James Stuart Olson and Mariah Gumpert, "National Defense Education Act of 1958," in *The 1950s: Key Themes and Documents* (Santa Barbara, CA: ABC-CLIO, 2018), Gale eBooks.

¹⁸⁷ Sean Kay, "America's Sputnik Moments.," *Survival* (0039-6338) 55, no. 2 (2013): 126, <https://doi.org/10.1080/00396338.2013.784470>.

¹⁸⁸ "The Changing Curriculum," in *American Decades*, ed. Judith S. Baughman, et al. (Detroit, MI: Gale, 2001), 7: Gale in Context: U.S. History.

¹⁸⁹ Michael S. Neiberg, "GI Bill of Rights," in *World War II: The Definitive Encyclopedia and Document Collection*, ed. Spencer C. Tucker (Santa Barbara, CA: ABC-CLIO, 2016), 2:Gale eBooks.

¹⁹⁰ Kay, "America's Sputnik," 4.

was taken aback more by the “intensity of public concern.”¹⁹¹ Immediately after Sputnik, Eisenhower and his advisors began to develop a response to the public. Eisenhower believed that the American people needed to remain informed about the administration’s policies, so that they would support what was necessary and demonstrate continued interest in the Space Race. He formulated an information campaign with the goal of supplying the public with facts in an accessible manner that stressed American technological capabilities. To accomplish this, he intentionally limited the scope of his public addresses to a few key points. In November, 1957, this strategy was visible during two of his speeches to the public. Eisenhower specifically emphasized American military strength and outlined upcoming policy alterations like expanded missile research, acceleration of space programs, increased defense spending, and a dispersal of the Strategic Air Command.¹⁹² Sputnik also forced Eisenhower to confront the need for more robust weapons and research programs. However, considering America’s fiscal constraints and the public’s panic, Eisenhower concluded that it would be most effective to shift program priorities based on their potential to win public support. As a result, under Eisenhower’s supervision, more than 5% of the defense budget (\$2B annually) was redirected toward publicly favored programs. The administration deemed intermediate range ballistic missiles, for instance, beneficial to boosting public morale, so they became a defense research and development priority.¹⁹³ Eisenhower also pushed for acceleration of satellite and space programs, refusing to allow any delays for the Vanguard satellite launch scheduled for December 6th, despite warnings from the Navy that the program was not ready. American prestige was further damaged when Vanguard proceeded to explode shortly after its launch. Following this failure, the Army’s Orbiter program made a second attempt to reach space with its Explorer satellite. Explorer-1 successfully launched from Cape Canaveral, Florida, on January 31st, 1958, reassuring American audiences that the U.S. was not as far behind the Soviets as many assumed. Nonetheless, the launch of Explorer-1 garnered only a fraction of the international press coverage and excitement of the Soviet launch of Sputnik-1.¹⁹⁴

The anemic start led the Eisenhower administration to redouble its efforts by eliminating bureaucratic impediments to progress. One obstacle, according to officials like former Assistant Secretary of the Air Force for R&D Trevor Gardner, was interservice rivalry within the military. Gardner explained that “[there were] at least nine ballistic-missile programs, all competing for roughly the same kind of facilities, the same kind of expertise, the same kind of engines, and the same public attention.”¹⁹⁵ Rather than focusing the government’s energy into one impressive rocket program that could triumph over Russia, the sectors of the military were caught up in battling one another for their own R&D funds.¹⁹⁶ Unfortunately, this interservice conflict was neither a novel nor isolated occurrence. It was a more dramatic episode of a long-standing

¹⁹¹ Payne, "Public Opinion," Page 95.

¹⁹² Payne, "Public Opinion," Page 96-97.

¹⁹³ Payne, "Public Opinion," 99.

¹⁹⁴ Lawrence, "The Space."

¹⁹⁵ "The Race."

¹⁹⁶ Walter A. McDougall, *The Heavens and the Earth: A Political History of the Space Age* (New York: Basic Book, 1985), Page 79.

problem. The advent of atomic weaponry in 1945 had led to bickering and strife, between the Army and Navy services. They argued over their respective missions and roles. The creation of the Air Force in 1947 reduced but did not alleviate these tensions. The 1949 struggle for control over strategic aviation between the Navy and Air Force proved their continuation.¹⁹⁷

To streamline satellite production, Eisenhower reorganized the Department of Defense and centralized military R&D funding. In July of 1958, Congress passed the National Aeronautics and Space Act, leading to the formation of the National Aeronautics and Space Administration, more commonly known as NASA. In addition, the Space Act established an advisory council to guide Eisenhower in space exploration, the Advanced Research Projects Agency to oversee military space activities, and a civilian-military liaison committee to coordinate NASA and DoD initiatives.¹⁹⁸ Later, in October 1959, Eisenhower shifted Von Braun and his team from the Army's space program to NASA. Considering that these scientists comprised 4,400 of the program's 5,500 developers, Eisenhower effectively ended the Army's first space program.¹⁹⁹

At the same time that it became the focus of government attention, the Space Race also permeated American popular culture in everything from toys to television. After the launch of Sputnik, American toymakers wasted no time, immediately designing Sputnik-inspired toys for the thousands of newly space-obsessed children. In an issue of *TIME* published on October 21st, 1957, the magazine's "SELLING" section featured an image of Ideal Toys' new product: the "Satellite Launcher". The accompanying article noted that within seven days, Ideal Toys had sold over 100,000 units of the Satellite Launchers. They were each sold for \$4.98 and included a rotating radar tracking station, as well as four plastic disks that could be fired 75-feet into the air. Ideal Toys also sold 50,000 units of its "Sky Sweeper Trucks" which projected searchlight silhouettes of jet planes against walls and shot these "planes" down with two "Nike" rockets. Acknowledging that Sputnik symbolized a geopolitical and technological threat to the U.S., an Ideal Toy executive still expressed gratitude for the satellite due to the business it attracted for the toy industry, "This may be a propaganda blow to the U.S.. But for us--boy, oh, boy."²⁰⁰ Other companies similarly produced a variety of innovative space toys including a parachute rocket powered by water and air pressure, a two-stage lunar rocket that required "harmless" fuel to travel three hundred feet, and a battery-powered "count-down" rocket-missile launcher.²⁰¹

Television programming also became dominated by all things cosmic. About a year after Sputnik-1, in September of 1958, the *Christian Science Monitor* reported that science-fiction dramas were rapidly appearing across all TV channels. The article cited "Mr. Krane" as an example. The one hour drama depicted an alien from the Milky Way as it traveled to Earth to

¹⁹⁷ Michael J. Neufeld, "The End of the Army Space Program: Interservice Rivalry and the Transfer of the Von Braun Group to NASA, 1958-1959," *The Journal of Military History* 69, no. 3 (2005): 738, JSTOR Journals.

¹⁹⁸ Lawrence, "The Space."

¹⁹⁹ Neufeld, "The End of the Army," 739.

²⁰⁰ "Into the Orbit," *TIME Magazine* 70, no. 17 (1957): Time Magazine Archive.

²⁰¹ Gay Pauley, "Orbit, Old West Toys Favored: Old West, Space Toys Favored," *Los Angeles Times (1923-1995)* (Los Angeles, Calif.), July 17, 1959, ProQuest Historical Newspapers: Los Angeles Times (167500666).

prevent humans from destroying the planet with warfare.²⁰² “The Blob” was another horror, sci-fi movie that showcased an alien blob that landed on Earth and engulfed anything in its path. These movies are only two examples of the many space themed television shows that were aired throughout the 1960s. In fact, according to Internet Movie Database, the 1960s saw the production of twenty-seven space-centered shows. These included famous programs like *Lost in Space* and *Star Trek*.²⁰³

With the press, toy, and television industries’ encouragement, it was no surprise that the American youth became enamored with outer space. It became customary for teenage boys to form rocketry clubs in their backyards and build their own small-scale rockets using magazine tutorials and their own wild imaginations. In *Rocket Boys*, Homer Hickam, who would go on to be a NASA engineer, details his experience as a high school boy during Sputnik. He and his friends would carefully follow the instructions of a *Life Magazine* article that delineated the process of building one’s own miniature rocket. While in their first stages of “rocket development”, the boys were not deterred by their lack of resources, instead creatively employing products like cherry powder to use as rocket fuel. Their persistence earned them stands at prestigious science fairs and the praise of their small-town neighbors.²⁰⁴

SECTION III: THE APOLLO PROJECT

Between 1958 and 1961, the United States continued to struggle in the Space Race against the Soviets. The USSR consistently and rapidly succeeded in their innovations throughout the late fifties and early sixties. Although the launch of Explorer I was viewed as a milestone for the American space program, it was quickly surpassed. In September 1959 the Soviets landed the unmanned spacecraft Lunik II on the moon. A month later, they landed a second craft, Lunik III, that was able to transmit the first images of the moon’s “dark side” back to Earth. Less than a year later, the Kremlin launched the 3,000 pound Sputnik-III, which sent two dogs, four mice, one rat, a jar of flies, algae, and fungi into orbit. The Soviets received these creatures back on Earth the following day, asserting their ability to safely send living beings into space. By contrast, American achievements during this period were far more modest. In the late 1950s, NASA succeeded in launching a series of small satellites into orbit, one of which, TIROS-1, was able to photograph weather patterns.²⁰⁵

The comparative weakness of the U.S. space program led to widespread concerns both at home and abroad about the level of American technological and military prowess. Foreign powers questioned whether American failures were evidence that the U.S. was not capable of protecting the western Europe from a nuclear attack. In October 1960, *The New York Times* and *Washington Post* leaked confidential polls conducted by the United States Information Agency

²⁰² M.M, "Quiz Show and Fantasy into Orbit: Radio-TV," *The Christian Science Monitor* (1908-) (Boston, Mass.), September 16, 1958, ProQuest Historical Newspapers: International Collection (509796337).

²⁰³ Mike Sean, "Sci-Fi Television in the 1960s," Internet Movie Database, last modified 2015, accessed April 17, 2025, <https://www.imdb.com/list/ls000097399/>.

²⁰⁴ Hickam, *Rocket Boys*, Page 36.

²⁰⁵ Bradley G. Shreve, "THE US, THE USSR, AND SPACE EXPLORATION, 1957-1963," *International Journal on World Peace* 20, no. 2 (2003): 79, JSTOR.

(USIA). The polls painfully showcased America's worsening global reputation to American and international audiences. *Washington Post* correspondent Warner Unna reported that within five major Western European nations (Britain, France, West Germany, Italy, and Norway), no more than 13% of those polled believed that the United States was ahead of the Soviet Union in rocketry. Unna went on to note that foreign poll respondents equated Soviet space and military capabilities, believing that the "USSR [would] emerge [as] the militarily stronger power over the next decades..."²⁰⁶ *The Times* discussed USIA polls specific to Britain and France. In Britain, 51% of voters answered that they had "not very much" or "very little" faith in the US to lead the West "in dealing with present world problems". In France, an even greater 62% responded that they had little confidence in the US.²⁰⁷ Reports such as these presumably reinforced the American public's feelings of humiliation, fear, and frustration among the American public. In particular, the public was likely worried that Soviet advancements in space meant they were less secure at home. News of a CIA disaster in Cuba in April 1961 did nothing to allay concerns.

The Bay of Pigs fiasco could not have come at a worse time for the United States. On April 12th, 1961, the USSR broke the news of its biggest achievement yet. Soviet cosmonaut Yuri Gagarin had just safely returned from space after a successful eighty-nine minute orbit of the earth.²⁰⁸ The first human had officially completed a trip into space, and he was a Russian. Once again, the Soviet Union had outdone the United States. Five days later, as hundreds of international news outlets exploded with headlines covering Gagarin's journey, news also broke that a CIA operation to remove Fidel Castro had failed, resulting in the deaths of nearly 300 CIA trained and led Cuban rebels, and the capture of nearly 1,200 more.²⁰⁹ The story of Gagarin's flight coupled with news of the CIA failure in Cuba served as the ultimate Communist propaganda news package. Media outlets painted the shameful picture: while the Soviets were courageously moving humanity forwards in the pursuit of scientific knowledge, the United States was engaged in a dishonest and destructive foreign policy against a weak neighboring country.

Twenty-seven-year-old Gagarin quickly became a worldwide sensation and a symbol of Soviet greatness throughout the remainder of 1961. The *Los Angeles Times* detailed his trip to Britain in July 1961, sharing that Gagarin faced mobs of fans in Manchester, lunched with the Queen in Buckingham Palace, and even delivered inspirational speeches to thousands of young British children.²¹⁰ In a similar visit to India in December 1961, Gagarin was invited to address the students of Lucknow University. He kindly encouraged their interests in outer space, "I am

²⁰⁶ By Warren Unna Staff Reporter, "U.S. Prestige Slip Seen in Poll of 5 Major Allies: Based on Polls," *The Washington Post, Times Herald* (1959-1973) (Washington, D.C.), October 29, 1960, ProQuest Historical Newspapers: The Washington Post (141200431).

²⁰⁷ "PRESTIGE LOSSES by U.S. AND SOVIET SHOWN in SURVEY: U.S.I.A. STUDY COVERS POLL in BRITAIN AND FRANCE -- SECRECY POLICY BACKED U.S. SURVEY CITES LOSS in STANDING," *New York Times* (1923-) (New York, N.Y.), October 27, 1960, ProQuest Historical Newspapers: The New York Times (115156737).

²⁰⁸ Shreve, "THE US, THE USSR," 79..

²⁰⁹ Joseph Siracusa, "Bay of Pigs," in *America in the World, 1776 to the Present: A Supplement to the Dictionary of American History*, ed. Edward J. Blum (Farmington Hills, MI: Charles Scribner's Sons, 2016), 1: Gale eBooks.

²¹⁰ "Yuri Gagarin Has Luncheon with Queen," *Los Angeles Times* (1923-1995) (Los Angeles, Calif.), July 15, 1961, ProQuest Historical Newspapers: Los Angeles Times (167914059).

told you have a space club. I am sure representatives of your university will go to space one day.”²¹¹ Meanwhile, in Washington, the mood was less celebratory. Responsibility for the Bay of Pigs fiasco fell squarely on President John F. Kennedy who “took the failure personally”.²¹² Theodore Sorensen, Kennedy’s speechwriter and a close advisor, recalled that Kennedy was “depressed and lonely” during this period. He noted that it was as if Kennedy had “handed his critics a stick with which they would forever beat him.”²¹³ As Karsten Werth explained it, Kennedy’s failure in Cuba painted him as incompetent in world affairs, and it even implicated the American people as it seemed that they had voted for a president who was unqualified.²¹⁴ In addition, the Bay of Pigs occurring alongside Gagarin’s flight seemed to confirm that America’s inferiority in space was not a simple ruse; instead, America’s shortcomings in space seemed to accurately reflect its declining abilities in realms even beyond technology.²¹⁵

The chaos of April 1961 and the Soviets’ tremendous win had detrimental effects on the American public’s morale. Marilyn C. Trainer of Silver Spring, Maryland, wrote to the editors of the *Washington Post*, describing her inability to “avoid a feeling of profound sadness because this piece of history should have been, and could have been [America’s].”²¹⁶ NASA’s fledgling astronaut Virgil Grissom voiced similar feelings of disappointment, “Obviously I’m disappointed that we weren’t first,” and Bem Price of the *Seattle Daily Times* characterized Gagarin’s orbit as yet another blow to America’s “scientific pride”.²¹⁷ Even NASA officials articulated feelings of hopelessness in interviews, claiming that America had lost the Space Race years earlier in 1954 when the Soviets began their space efforts and America did not.²¹⁸ A frustrated Hanson W. Baldwin, a *New York Times* military specialist, responded to Gagarin with a critique targeted at the Kennedy administration. In his article titled “Flaw in Space Policy...”, Baldwin demanded more “Presidential emphasis” and urgency be directed towards the Space Race. He declared that the Eisenhower-era philosophy of prioritizing scientific achievement over “spectacular” missions had to be discarded if the U.S. wanted to reclaim its prestigious image, “Now that the Russians have capped their space ‘firsts’ with the tremendous feat of putting a man into space, it is high time to discard this policy.” Baldwin’s article suggested that the United States was not actually

²¹¹ "Yuri Gagarin in Lucknow: WARM WELCOME ACCORDED," *The Times of India (1861-)* (Mumbai, India), December 2, 1961, ProQuest Historical Newspapers: International Collection (613289663).

²¹² Siracusa, "Bay of Pigs,".

²¹³ John M. Logsdon, *The Decision to Go to the Moon: Project Apollo and the National Interest* (Cambridge, MA: MIT Press, 1970), 111.

²¹⁴ Werth, "A Surrogate,"571.

²¹⁵ John M. Logsdon, *The Decision to Go to the Moon: Project Apollo and the National Interest* (Cambridge, MA: MIT Press, 1970), 112.

²¹⁶ Alfred Sherrard et al., "Letters to the Editor: Fallout on Snow," *The Washington Post, Times Herald (1959-1973)* (Washington, D.C.), April 23, 1961, ProQuest Historical Newspapers: The Washington Post (141304458).

²¹⁷ *Seattle Daily Times* (Seattle, Washington), April 12, 1961, America's Historical Newspapers.

²¹⁸ H. R. Vohra, "Wave of Self-Criticism Sweeps U.S.: DEFEAT in SPACE RACE ACCEPTED SPORTINGLY the Times of India News Service," *The Times of India (1861-)* (Mumbai, India), April 15, 1961, ProQuest Historical Newspapers: International Collection (741097163).

inferior in science, having collected valuable data in space, but that its failure to capitalize on the propaganda value of the missions had caused its international reputation to falter.²¹⁹

American morale and the Kennedy administration's reputation was buoyed briefly on May 5th, 1961, when Navy pilot Alan Shepard became the first American in space with his Freedom 7 sub-orbital flight.²²⁰ A proud nation cheered his success. A New Hampshire highway was named after Shepard, and he even received a medal from President Kennedy before a crowd of more than two hundred fifty thousand. However, Shepard failed to attain the level of international praise that Gagarin had as only national news outlets reported on his orbit.²²¹ This left Kennedy with the understanding that more and greater successes were needed.

On May 25th, 1961, Kennedy addressed a joint session of Congress. His address, formally titled "Special Message to the Congress on Urgent National Needs", would go down in history as the speech that initiated the greatest space mission known to man: the moon landing. Kennedy proudly proclaimed that by the end of the decade, America should send a man to the moon, "I believe that this nation should commit itself to achieving the goal, before this decade is out, of landing a man on the moon and returning him safely to the Earth. No other space project in this period will be more impressive to mankind."²²² It was a remarkable challenge. The moon was a quarter million miles away and any journey to reach it would have to last a minimum of a week. So far, the U.S. had only managed to send a man into space for 15 minutes on a journey of 116 miles.²²³ Nonetheless, Kennedy wished to redeem America's image as the world's premiere technological and innovative power. He believed that in order to do so, the United States had to grander than any being explored by the USSR. His hope was that a massive success in space would avert public attention away from recent mistakes, and inspire pride across the world. Furthermore, he believed the project would ignite a sense of national unity, "it will not be one man going to the moon--if we make this judgement affirmatively, it will be an entire nation."²²⁴ In this way, Kennedy sought to extend the Apollo Project to include the public; he made average Americans feel that by backing the Apollo Project, they too were playing a vital role in defending American interests in the Cold War. Resultantly, Americans were more invested in the Space Race.

Following his address, Gallup polls and news articles showed strong evidence of increasing public morale. In one poll published on June 4th, 1961, just days after Kennedy's

²¹⁹ By HANSON W. BALDWIN, "Flaw in Space Policy: U.S. Is Said to Lack Sense of Urgency in Drive for New Scientific Conquests," *New York Times* (1923-) (New York, N.Y.), April 17, 1961, ProQuest Historical Newspapers: The New York Times (115440504).

²²⁰ Lawrence, "The Space."

²²¹ Philip Dodd, "250,000 Hail Shepard in Capital: HE RECEIVES a MEDAL AND MEETS PRESS Astronaut Goes Back to Base Spaceman Gets Hero Medal and Capital Greeting," *Chicago Daily Tribune* (1923-1963) (Chicago, Ill.), May 9, 1961, ProQuest Historical Newspapers: Chicago Tribune (182917131); "U.S. Highway Named for Alan Shepard," *The Atlanta Constitution* (1946-) (Atlanta, Ga.), June 30, 1961, ProQuest Historical Newspapers: The Atlanta Constitution (1616076818).

²²² "John F. Kennedy," in *Space Exploration Reference Library*, ed. Lawrence W. Baker, et al. (Detroit, MI: UXL, 2005), 4, Gale in Context: U.S. History.

²²³ A. Bowdoin Van, "Project Apollo," in *Encyclopedia of the Sixties: A Decade of Culture and Counterculture*, ed. James S. Baugess and Abbe Allen DeBolt (Santa Barbara, CA: Greenwood, 2012), 2, Gale eBooks.

²²⁴ "John F. Kennedy," 4.

historic announcement, Americans were asked “Which country--the United States or Russia--do you think will be the first to send a man to the moon?” The results showed a profound change from months earlier when a large majority of Americans believed the Soviets would be first. After Kennedy’s speech, the number of Americans who believed the Soviets would be first plummeted to only 34%, while 33% now believed the US would be first. The final 33% held no opinion.²²⁵ Newspaper columnists reflected the newfound enthusiasm as well. In July 1961 writer for the *Atlanta Constitution*, invoked the wisdom of fables and compared the contest between the U.S. and USSR to the race between the tortoise and the hare. While the Soviet Union may have arrogantly leapt ahead in the race’s early stages, in the end, the persistence and strategy of the U.S. would win the day like the tortoise.²²⁶

Americans’ growing inclination to support the Apollo Project was heightened by the appearance and public demeanor of the seven astronauts who were chosen for the program. In alignment with Kennedy’s vision of American unity and pride, the men picked were meant to embody traditional American values and relate to common U.S. citizens. Hence, Americans felt an affinity for the moon-landing program as they recognized parts of themselves in the astronauts: athletic, clean-cut, and conservatively dressed young men who brimmed with patriotism and a desire to win the Cold War for America.²²⁷ John Glenn, who became the first American to orbit the Earth in 1962, was a prime example. Glenn had been born to a Presbyterian Family in the small town of New Concord, Ohio. He had grown up on Friendship Drive and married his high school sweetheart, who was also his neighbor. He did not drink or smoke.²²⁸ In an April 1962 issue of the *Chicago Tribune*, Saul Pett described Glenn as a “thoroughly likeable guy” with an “uncommon common touch.” Pett highlighted the affection and personal identification that the many Americans felt towards Glenn, atypical of past political figures. Glenn, as well as other early astronauts like Virgil Grissom and Alan Shepard, became national heroes who galvanized the public. As Pett put it, “[Glenn made] Americans feel better about being Americans. He did start a comeback in national morale, which...had a long way to go.”²²⁹ In addition, beyond their lauded backgrounds and personalities, astronauts like Glenn were able to rally the American public due to their appearances. The charming smiles, freckled faces, and masculine figures of NASA’s astronauts reflected the familiar ones quintessential of American fathers, husbands, and friends, pulling on the heartstrings of American audiences.²³⁰

American support of the Apollo Project was also apparent in the intense commitment of NASA employees. Apollo dominated the lives of many NASA scientists who spent years working—often in isolation—with the sole focus of getting a man on the moon. They typically

²²⁵ George Gallup, "GALLUP POLL: SPACE RACE NOW SEEN as EVEN," *Los Angeles Times* (1923-1995) (Los Angeles, Calif.), June 4, 1961, ProQuest Historical Newspapers: Los Angeles Times (167870965).

²²⁶ "Tale of Hare, Tortoise Can Apply in Space Race," *The Atlanta Constitution* (1946-) (Atlanta, Ga.), July 13, 1961, ProQuest Historical Newspapers: The Atlanta Constitution (1616077015).

²²⁷ Van, "Project Apollo," 2.

²²⁸ Matt Bai, "The Time Traveler," *Newsweek*, October 26, 1998, Gale in Context: U.S. History.

²²⁹ Saul Pett, "Time, Man, Event Shape an American Hero: Col. John Glenn: Writer Looks at Making of Astronaut," *Chicago Daily Tribune* (1923-1963) (Chicago, Ill.), April 8, 1962, ProQuest Historical Newspapers: Chicago Tribune (183149897).

²³⁰ The Granger Collection, *John Glenn*, photograph, February 20, 1962, Britannica ImageQuest.

worked more than fifteen hours per day , and were often so focused on Apollo that they were unaware of any other world events. In Charles Murray and Catherine Bly Cox's "Apollo: The Race to the Moon", NASA employees are quoted sharing their extreme devotion to Apollo. One employee stated that they had "missed the entire Vietnam War," while another admitted that he "watched no television, read no newspapers, came to work at six in the morning, and worked until nightfall, six or seven days a week for years." Former NASA flight director Christopher C. Kraft recounted his commitment and that of his colleagues in his "Flight: My Life in Mission Control", "We knew that, come hell or high water, we were going to put an American into orbit around the world...the future of our American way of life was at stake."²³¹ These employees believed that even their smaller contributions towards the Apollo effort could influence the greater Cold War, protecting their democracy from communism. They exemplified the success of Kennedy's plan, uniting patriotically against the threat of the Soviet Union.

Sadly, the optimism and enthusiasm generated by the Apollo Program did not survive the 1960's. On November 22nd, 1963, John F. Kennedy was assassinated as he was paraded down the streets of Dallas, Texas. The assassination profoundly shook the American public who had grown accustomed to Kennedy's frequent television addresses and when his sincere, patriotic words echoed through their living rooms. Kennedy had led America into the next decade of the Cold War with dignity and had pushed the nation to embark on the awesome mission to the moon. Kennedy's funeral rites, which were televised on all of the nation's networks, amassed a viewership of 93% of television-possessing households in the United States. Though the Apollo Project was viewed as a memorial for the slain president, it lacked its courageous leader.²³² To officially pay tribute to JFK's efforts in space, America's new President Lyndon B. Johnson appeared to the public in his first broadcast on November 29th, announcing that Cape Canaveral would be renamed Cape Kennedy.²³³ He attempted to encourage the saddened nation to look onwards and rise above the tragedy, "A great leader is dead; a great nation must move on." Even more, Johnson asked the public to unite as Kennedy had desired, discarding their bias, "It is the work that I most want us to do, to...protect our unity north and south, east and west; to hasten the day when bias of race, religion, and region is no more."²³⁴ Unfortunately, as the events of the 1960's unfolded, Johnson's vision would not be realized.

With the passing of the Gulf of Tonkin Resolution in 1964, United States troops became active participants in the Vietnam War. While most Americans initially supported American intervention in the war, as the conflict escalated and the death toll mounted, support began to

²³¹ Werth, "A Surrogate,"576.

²³² "Assassination and Violent Protest," in *American Decades*, ed. Judith S. Baughman, et al. (Detroit, MI: Gale, 2001), 7 Gale in Context: U.S. History.

²³³ By Carroll Kilpatrick Staff Reporter, "Canaveral Now Cape Kennedy: Johnson Renames Space Hub President Urges Unity in Nation, End of Rancor Witnessed Polaris Launch Emphasizes Future Johnson Renames Canaveral, Space Center to Honor Kennedy," *The Washington Post, Times Herald (1959-1973)* (Washington, D.C.), November 29, 1963, ProQuest Historical Newspapers: The Washington Post (141897369).

²³⁴ Lyndon B. Johnson, "State of the Union Address: Lyndon B. Johnson, January 8, 1964.," *History Reference Ultimate*, last modified March 1, 2006, <https://research.ebsco.com/linkprocessor/plink?id=e013c0b9-ed6c-3a3b-bf46-df472b202d0f>.

flag. South Vietnam's victory grew less plausible due to the overwhelming success of North Vietnam's guerilla warfare, and the public began equating shipping Americans to Vietnam with shipping them off to their death. Frequent news stories about the endless bombing of the North and the use of the herbicide Agent Orange continued to inflame public opinion. As years passed and Johnson refused to concede to the public and withdraw American troops, confidence in the government declined precipitously.²³⁵

Vietnam was not the only issue that divided Americans in the 1960's. Urban riots over the issues of poverty and race rocked nearly every region. In 1960, the Conference on Economic Progress estimated that over 77 million Americans lived in poverty, amounting to more than two fifths of the whole nation.²³⁶ Prior to his assassination, Kennedy had begun researching legislation to remedy the nation's poverty. Although his death interrupted those efforts, Johnson was adamant to succeed in fulfilling Kennedy's vision of social justice. In his 1964 State of the Union Address, Johnson boldly declared a "war on poverty", "This administration today, here and now, declares unconditional war on poverty in America." In August 1964, he signed the Equal Opportunity Act, initiating a robust social reform program that focused on breaking the cycle of poverty for millions of Americans. The act established funding for a variety of initiatives: Head Start for preschool children, Job Corps to provide assistance in vocational training, work-study programs, grants for businesses and farmers open to hiring unemployed Americans, community action programs, and college admissions aid. Between 1965 and 1966, Congress passed over one hundred bills dedicated to improving quality of life in America. Lamentably, many of these plans ended in the face of the fiscal demands of the Vietnam War and the Space Race.²³⁷ As a result, Johnson's "war on poverty" did not result in as great of change as promised, prompting anger amongst Americans across the nation. Struggling Americans felt abandoned.

The political chaos of the 1960s set the stage on which the Apollo Project came to fruition. For some Americans, the Apollo Project served as the ultimate symbol of American greatness amidst an otherwise bleak landscape. Apollo provided a source of unity when the nation lacked it most. Others, however, questioned how the government could justify pouring time and funding into Space when there were so many other obvious challenges closer to home. To this group of Americans, the moon landing was evidence that the government was capable of investing greater resources into solutions to the problems of poverty and racism, but simply chose not to.

On July 20th, 1969, Neil Armstrong and "Buzz" Aldrin landed on the moon, officially bringing JFK's Apollo Project to fruition with Apollo 11. The two astronauts arrived on the moon's surface, touching down in lunar module Eagle, named for America's national bird. As an

²³⁵ "Vietnam War," in *Gale U.S. History Online Collection* (Detroit, MI: Gale, 2024), Gale in Context: U.S. History.

²³⁶ KERRY F. NAPUK. Arlington et al., "Letters to the Editor: Attack on Poverty," *The Washington Post, Times Herald (1959-1973)* (Washington, D.C.), January 10, 1964, ProQuest Historical Newspapers: The Washington Post (142364698).

²³⁷ "Economic Opportunity Act," in *Gale Encyclopedia of U.S. Economic History*, ed. Thomas Carson and Mary Bonk (Detroit, MI: Gale, 1999), [Page #], Gale in Context: U.S. History.

estimated half a billion people watched on live television, Armstrong walked the lunar plain known as the Sea of Tranquility, uttering his famous words, “one small step for [a] man, one giant leap for mankind.”²³⁸ Around the world, people were overwhelmed with indescribable joy. Walter Cronkite, the news reporter who had broken the news of JFK’s death six years earlier, excitedly exclaimed “Armstrong is on the moon! Neil Armstrong, a 38-year-old American, standing on the surface of the moon!”²³⁹ A July 21st *Washington Post* article detailed the celebrations that occurred across the globe. In London, there were screams of joy resounding through Trafalgar Square; in Chile, people danced in the streets; hundreds of Poles crowded the US embassy in Warsaw; and in the Middle East, broadcasts updating listeners on the War of Attrition were disrupted to announce the landing.²⁴⁰ In particular, a reaction that made headlines was that of two Russian boys who were reported to have called out “Hooray!”²⁴¹ and “It’s a great day.” The pride in humankind that Apollo 11 inspired was unhampered by Cold War divisions, vast oceans, or national borders. This reality of Apollo had exceeded Kennedy’s expectations, bringing together populations beyond the United States.

As both international and national news outlets published hundreds of articles praising the United States’ magnificent feat, it was clear that American prestige had been restored. Nevertheless, a notable number of Americans articulated feelings of frustration after the project’s success. Letters written to the *Los Angeles Times* depict a mixed reaction. Within a sample of eight letters, two articulated pure excitement for the landing, another two voiced that they appreciated the accomplishment but now wished for equal success in achieving world peace, and four lamented that they couldn’t genuinely enjoy the landing when their nation was riddled with political and economic turmoil.²⁴²

Other newspapers also published distraught letters written by readers. *The Post* highlighted Marie McBride, a teacher working in an underserved school in Washington D.C. who harbored anger towards the American government. She explained that while the government was capable of making an “all-out national effort” to reach the moon, it failed to take similarly effective measures to provide for the impoverished children that she teaches everyday. She

²³⁸ By Thomas O'Toole Washington Post Staff Writer, "'One Small Step for Man ... Giant Leap for Mankind': American Flag Planted Amazingly Clear Picture Guides Aldrin down Ladder Armstrong, Aldrin Become First Men to 'Walk on Moon Bleak but Beautiful 'Every Variety of Shapes' 'Like Being in an Airplane' 'Just Play It Cool' Falling, Slowing Approach Couldn't Pinpoint Location Collins Wakes up First Activates Landing Craft Elaborate Instrument Check Apollo Telecast.," *The Washington Post, Times Herald (1959-1973)* (Washington, D.C.), July 21, 1969, [Page #], ProQuest Historical Newspapers: The Washington Post (147628777).

²³⁹ Samuel Willard Crompton, "President Nixon and the Space Race," in *Sputnik/Explorer I, Updated Edition* (Chelsea House, 2017), Modern World History.

²⁴⁰ By Robert C. Jensen Washington Post Staff Writer, "Millions around the World Hail Lunar Landing; Follow It on TV: Moon Landing Seen around World War News Stops 'With Our ... Prayers' Euphoria in Warsaw Half Billion Watch," *The Washington Post, Times Herald (1959-1973)* (Washington, D.C.), July 21, 1969, ProQuest Historical Newspapers: The Washington Post (147627571).

²⁴¹ "A Russian Yells 'Hooray': The World Celebrates as Armstrong, Aldrin Land Safely on the Moon," *The Globe and Mail (1936-)* (Toronto, Ont.), July 21, 1969, ProQuest Historical Newspapers: International Collection (1242409284).

²⁴² R. Nathan Baker et al., "LETTERS to THE TIMES: READERS VIEW APOLLO'S ACHIEVEMENTS in LIGHT of EARTHLY NEEDS, BENEFITS," *Los Angeles Times (1923-1995)* (Los Angeles, Calif.), July 24, 1969, ProQuest Historical Newspapers: Los Angeles Times (156308172).

testified that the “sons of Americans” were going without decent housing and food daily, but instead of improving their living conditions, the nation focused on collecting rocks from space.²⁴³ A majority of black Americans seemed to agree with McBride. Reverend C.W. Cummings bluntly stated, “The money should have been used to develop ghetto areas nationwide,” and auto factory worker Elijah Davis declared, “They’ve got no business up there. The money should be used to help the poor.”²⁴⁴ Journalist Art Seidenbaum also noted that during the advent of the Apollo Project, when America committed to reaching the moon, it was unaware of the struggles that the sixties would bring. American leaders did not know in 1961 that \$30B a year would be spent on Vietnam, nor did they know that Johnson’s “war on poverty” would fail to amount to anything beyond “too many pamphlets [with] too little consequence.”²⁴⁵

In retrospect, the Apollo Project’s legacy is one of both inspiration and dissonance. It demonstrated the power of American mobilization, proving that the United States could accomplish what once seemed unthinkable. But, the boundless ambition of reaching the moon violently clashed with the pressing necessity for justice and equity amongst struggling populations of Americans.

CONCLUSION

Looking back today, Americans may marvel at the speed with which the American space program managed to reach the moon, or feel nostalgic for the national cause and patriotism that the Space Race presented. However, beyond these initial feelings, they may also realize that the Space Race symbolized far more than a technological contest between two superpowers; it was a mirror reflecting America’s hopes and most pressing worries.

Born in the wake of World War II and the capture of German rocket technology, the race to space became a stage on which national identity, ideological superiority, morality, and geopolitical prestige were relentlessly performed. The launch of the Soviet’s Sputnik completely transformed American public policy, influencing the day-to-day lives of Americans through its invigoration of the education system and the imagination it lent to popular culture. Soon after, as the Apollo Project developed throughout the course of the 1960s and finally triumphed in 1969, Americans experienced a milestone in human history and a redemption of American prestige. Nonetheless, they also learned of the paradox in their national priorities. Even as half a billion people witnessed in awe as Neil Armstrong took man’s first steps on the moon, millions of Americans looked up at the stars with feelings of betrayal and skepticism, having been burdened by social unrest and poverty back on Earth. Entangling science, politics, and raw human reaction, the Space Race was a rare phenomenon that will continue to be studied for centuries to come.

²⁴³ Craig Colgate Jr., Jesse Silverglate, theos E. Manos, MARIE McBRIDE, “Letters to the Editor: Moon Vs. Children Ill-Will Ambassador Alsop on Vietnam NATO’s Southern Flank,” *The Washington Post, Times Herald* (1959-1973) (Washington, D.C.), October 1, 1969, ProQuest Historical Newspapers: The Washington Post (147627840).

²⁴⁴ “Billions for Trip to the Moon; Now How about Care for Poor,” *Michigan Chronicle* (1939-2010) (Detroit, Mich.), July 26, 1969, ProQuest Historical Newspapers: Black Newspaper Collection (2399456403).

²⁴⁵ Art Seidenbaum, “Afterthoughts on Apollo: The Moon Vs. Mankind?,” *Los Angeles Times* (1923-1995) (Los Angeles, Calif.), January 3, 1969, ProQuest Historical Newspapers: Los Angeles Times (156103802).

Works Cited

- "Assassination and Violent Protest." In *American Decades*, edited by Judith S. Baughman, Victor Bondi, Richard Layman, Tandy McConnell, and Vincent Tompkins. Vol. 7. Detroit, MI: Gale, 2001. Gale in Context: U.S. History.
- The Atlanta Constitution (1946-) (Atlanta, Ga.). "Tale of Hare, Tortoise Can Apply in Space Race." July 13, 1961, 4. ProQuest Historical Newspapers: The Atlanta Constitution (1616077015).
- The Atlanta Constitution (1946-) (Atlanta, Ga.). "U.S. Highway Named for Alan Shepard." June 30, 1961, 21. ProQuest Historical Newspapers: The Atlanta Constitution (1616076818).
- Bai, Matt. "The Time Traveler." *Newsweek*, October 26, 1998, 30. Gale in Context: U.S. History.
- Baker, R. Nathan, Terry Campbell, John A. Daniels, and HENRY D ALBERTS et al. "LETTERS to THE TIMES: READERS VIEW APOLLO'S ACHIEVEMENTS in LIGHT of EARTHLY NEEDS, BENEFITS." *Los Angeles Times* (1923-1995) (Los Angeles, Calif.), July 24, 1969, 1-b6. ProQuest Historical Newspapers: Los Angeles Times (156308172).
- Benson, Sonia, Daniel E. Brannen, Jr., and Rebecca Valentine. "Kellogg-Briand Pact." In *UXL Encyclopedia of U.S. History*, 849-50. Vol. 5. Detroit, MI: UXL, 2009. Gale in Context: U.S. History.
- By Carroll Kilpatrick Staff Reporter. "Canaveral Now Cape Kennedy: Johnson Renames Space Hub President Urges Unity in Nation, End of Rancor Witnessed Polaris Launch Emphasizes Future Johnson Renames Canaveral, Space Center to Honor Kennedy." *The Washington Post, Times Herald* (1959-1973) (Washington, D.C.), November 29, 1963, 2. ProQuest Historical Newspapers: The Washington Post (141897369).
- By HANSON W. BALDWIN. "Flaw in Space Policy: U.S. Is Said to Lack Sense of Urgency in Drive for New Scientific Conquests." *New York Times* (1923-) (New York, N.Y.), April 17, 1961, 5. ProQuest Historical Newspapers: The New York Times (115440504).
- By Institute of Public Opinion. "Polls Disclose U.S. Trend on Foreign Policy: Americans Now View Peace Conferences with Critical Eye." *The Washington Post* (1923-1954) (Washington, D.C.), June 27, 1937, 27. ProQuest Historical Newspapers: The Washington Post (150984585).
- By JOHN D. MORRIS Special to The New York Times. "Bigger Rocket an Urgent Need, Von Braun and Medaris Warn: Missile Experts Declare U.S. Will Lose Race to Russians Unless It Acts Fast STRONGER ROCKET HELD VITAL to U.S. Von Braun Makes a Hit Keller's Role Praised." *New York Times* (1923-) (New York, N.Y.), December 15, 1957, 1. ProQuest Historical Newspapers: The New York Times (114065439).
- By Robert C. Jensen Washington Post Staff Writer. "Millions around the World Hail Lunar Landing; Follow It on TV: Moon Landing Seen around World War News Stops 'With Our ... Prayers' Euphoria in Warsaw Half Billion Watch." *The Washington Post, Times Herald* (1959-1973) (Washington, D.C.), July 21, 1969, 2. ProQuest Historical Newspapers: The Washington Post (147627571).

By Thomas O'Toole Washington Post Staff Writer. "'One Small Step for Man ... Giant Leap for Mankind': American Flag Planted Amazingly Clear Picture Guides Aldrin down Ladder Armstrong, Aldrin Become First Men to 'Walk on Moon Bleak but Beautiful 'Every Variety of Shapes' 'Like Being in an Airplane' 'Just Play It Cool' Falling, Slowing Approach Couldn't Pinpoint Location Collins Wakes up First Activates Landing Craft Elaborate Instrument Check Apollo Telecast." *The Washington Post, Times Herald* (1959-1973) (Washington, D.C.), July 21, 1969, 2. ProQuest Historical Newspapers: The Washington Post (147628777).

By Warren Unna Staff Reporter. "U.S. Prestige Slip Seen in Poll of 5 Major Allies: Based on Polls." *The Washington Post, Times Herald* (1959-1973) (Washington, D.C.), October 29, 1960, 2. ProQuest Historical Newspapers: The Washington Post (141200431).

By Winzola McLendon. "Teen Scientists Have Answers: They're Space Age 'Pilots.'" *The Washington Post and Times Herald* (1954-1959) (Washington, D.C.), November 17, 1957, 2. ProQuest Historical Newspapers: The Washington Post (148839053).

"The Changing Curriculum." In *American Decades*, edited by Judith S. Baughman, Victor Bondi, Richard Layman, Tandy McConnell, and Vincent Tompkins. Vol. 7. Detroit, MI: Gale, 2001. Gale in Context: U.S. History.

Crompton, Samuel Willard. "President Nixon and the Space Race." In *Sputnik/Explorer I*, Updated Edition. Chelsea House, 2017. Modern World History.

Daily Defender (Daily Edition) (1956-1960) (Chicago, Ill.). "Says U. S. Lags in Space Race: Says U. S. behind in Space Race." November 25, 1957, 1. ProQuest Historical Newspapers: Black Newspaper Collection (493595615).

Delaney, David G. "Neutrality Acts." In *Major Acts of Congress*, edited by Brian K. Landsberg, 66-69. Vol. 3. New York, NY: Macmillan Reference USA, 2004. Gale in Context: U.S. History.

Dodd, Philip. "250,000 Hail Shepard in Capital: HE RECEIVES a MEDAL AND MEETS PRESS Astronaut Goes Back to Base Spaceman Gets Hero Medal and Capital Greeting." *Chicago Daily Tribune* (1923-1963) (Chicago, Ill.), May 9, 1961, 1. ProQuest Historical Newspapers: Chicago Tribune (182917131).

Dornberger, Walter R. "The German V-2." *Technology and Culture* 4, no. 4 (1963): 393-409. <https://doi.org/10.2307/3101376>.

"Economic Opportunity Act." In *Gale Encyclopedia of U.S. Economic History*, edited by Thomas Carson and Mary Bonk. Detroit, MI: Gale, 1999. Gale in Context: U.S. History.

The Editors of Encyclopaedia Britannica. "V-2 Rocket." *Britannica*. Last modified May 2, 2025. Accessed May 22, 2025. <https://www.britannica.com/technology/V-2-rocket>.

"Foreign Policy and the Road to War." *The Holocaust Explained*. Accessed May 22, 2025. [https://www.theholocaustexplained.org/life-in-nazi-occupied-europe/foreign-policy-and-the-road-to-war/build-up-to-war/#:~:text=Conscription%20and%20Rearmament%201935,the%20Luftwaffe%20had%208000%20planes](https://www.theholocaustexplained.org/life-in-nazi-occupied-europe/foreign-policy-and-the-road-to-war/build-up-to-war/#:~:text=Conscription%20and%20Rearmament%201935,the%20Luftwaffe%20had%208000%20planes.).

- Gallup, George. "GALLUP POLL: SPACE RACE NOW SEEN as EVEN." *Los Angeles Times* (1923-1995) (Los Angeles, Calif.), June 4, 1961, 1-i1. ProQuest Historical Newspapers: Los Angeles Times (167870965).
- The Globe and Mail (1936-) (Toronto, Ont.). "A Russian Yells 'Hooray': The World Celebrates as Armstrong, Aldrin Land Safely on the Moon." July 21, 1969, 9. ProQuest Historical Newspapers: International Collection (1242409284).
- The Granger Collection. John Glenn. Photograph. February 20, 1962. Britannica ImageQuest.
- Hickam, Homer H., Jr. *Rocket Boys*. New York: Dell Publishing, 1998.
- Hudson, David E. "DISARMAMENT URGED as AN ECONOMY MOVE: THE UNITED STATES IS VIEWED as HAVING SPLENDID OPPORTUNITY to save SAFELY." *New York Times* (1923-) (New York, N.Y.), January 22, 1933, 1. ProQuest Historical Newspapers: The New York Times (100599886).
- "Into the Orbit." *TIME Magazine* 70, no. 17 (1957): 94-97. Time Magazine Archive
- "John F. Kennedy." In *Space Exploration Reference Library*, edited by Lawrence W. Baker, Sarah Hermesen, Rob Nagel, and Peggy Saari, 50-59. Vol. 4. Detroit, MI: UXL, 2005. Gale in Context: U.S. History.
- Johnson, Lyndon B. "State of the Union Address: Lyndon B. Johnson, January 8, 1964." *History Reference Ultimate*. Last modified March 1, 2006.
<https://research.ebsco.com/linkprocessor/plink?id=e013c0b9-ed6c-3a3b-bf46-df472b202d0f>.
- Johnston, Robert. "U.S. Expenditures for Defense and Education, 1910-2021." Table. Johnston's Archive, January 29, 2023. Accessed May 22, 2025.
<https://www.johnstonsarchive.net/policy/edgraph.html>.
- Kay, Sean. "America's Sputnik Moments." *Survival* (0039-6338) 55, no. 2 (2013): 123-46.
<https://doi.org/10.1080/00396338.2013.784470>.
- KERRY F. NAPUK. Arlington, PO SUNG KIM. Arlington, van Der SLICE. Washington Jane K, CONVASS B. DEAN. Arlington, W.L. MOORE. Falls Church, Mich ZENOBIUS STELMACH. Highland Park, and Ala EDDIE H. LIVINGSTON. Dothan. "Letters to the Editor: Attack on Poverty." *The Washington Post, Times Herald* (1959-1973) (Washington, D.C.), January 10, 1964, 1. ProQuest Historical Newspapers: The Washington Post (142364698).
- Lawrence, Mark. "The Space Race and American Foreign Relations." In *The Oxford Encyclopedia of the History of American Foreign Relations*. Oxford University Press, 2021. Oxford Reference Online.
- Logsdon, John M. *The Decision to Go to the Moon: Project Apollo and the National Interest*. Cambridge, MA: MIT Press, 1970.
- . "Sputnik." In *America in the World, 1776 to the Present: A Supplement to the Dictionary of American History*, edited by Edward J. Blum, 970-72. Vol. 2. Farmington Hills, MI: Charles Scribner's Sons, 2016. Gale eBooks.

- Los Angeles Times (1923-1995) (Los Angeles, Calif.). "Yuri Gagarin Has Luncheon with Queen." July 15, 1961, 5. ProQuest Historical Newspapers: Los Angeles Times (167914059).
- MARIE McBRIDE, Craig Colgate Jr.jesse Silverglate.theos E. Manos. "Letters to the Editor: Moon Vs. Children Ill-Will Ambassador Alsop on Vietnam NATO's Southern Flank." The Washington Post, Times Herald (1959-1973) (Washington, D.C.), October 1, 1969, 1. ProQuest Historical Newspapers: The Washington Post (147627840).
- McDougall, Walter A. The Heavens and the Earth: A Political History of the Space Age. New York: Basic Book, 1985.
- Michigan Chronicle (1939-2010) (Detroit, Mich.). "Billions for Trip to the Moon; Now How about Care for the Poor." July 26, 1969, 8. ProQuest Historical Newspapers: Black Newspaper Collection (2399456403).
- "Missiles and Rockets: Sept. 7 - Dec. 1960." Internet Archive. Last modified January 8, 2024. Accessed May 23, 2025.
<https://archive.org/details/missilesrockets7196unse/page/n341/mode/2up>.
- Neiberg, Michael S. "GI Bill of Rights." In World War II: The Definitive Encyclopedia and Document Collection, edited by Spencer C. Tucker, 697-98. Vol. 2. Santa Barbara, CA: ABC-CLIO, 2016. Gale eBooks.
- Neufeld, Michael J. "The End of the Army Space Program: Interservice Rivalry and the Transfer of the Von Braun Group to NASA, 1958-1959." The Journal of Military History 69, no. 3 (2005): 737-57. JSTOR Journals.
- New York Times (1923-) (New York, N.Y.). "PRESTIGE LOSSES by U.S. AND SOVIET SHOWN in SURVEY: U.S.I.A. STUDY COVERS POLL in BRITAIN AND FRANCE -- SECRECY POLICY BACKED U.S. SURVEY CITES LOSS in STANDING." October 27, 1960, 1. ProQuest Historical Newspapers: The New York Times (115156737).
- Olson, James Stuart, and Mariah Gumpert. "National Defense Education Act of 1958." In The 1950s: Key Themes and Documents, 157-58. Santa Barbara, CA: ABC-CLIO, 2018. Gale eBooks.
- Pauley, Gay. "Orbit, Old West Toys Favored: Old West, Space Toys Favored." Los Angeles Times (1923-1995) (Los Angeles, Calif.), July 17, 1959, 2. ProQuest Historical Newspapers: Los Angeles Times (167500666).
- Payne, Rodger A. "Public Opinion and Foreign Threats: Eisenhower's Response to Sputnik." Armed Forces and Society 21, no. 1 (1994): 89-111. JSTOR.
- People, Henry. "Moon Landing Leaves Local Negroes Cold." Michigan Chronicle (1939-2010) (Detroit, Mich.), July 26, 1969, 1. ProQuest Historical Newspapers: Black Newspaper Collection (2399457728).
- Pett, Saul. "Time, Man, Event Shape an American Hero: Col. John Glenn: Writer Looks at Making of Astronaut." Chicago Daily Tribune (1923-1963) (Chicago, Ill.), April 8, 1962, 25. ProQuest Historical Newspapers: Chicago Tribune (183149897).
- "The Race to Come." TIME Magazine 70, no. 17 (1957): 21-23. Time Magazine Archive.

- Sean, Mike. "Sci-Fi Television in the 1960s." Internet Movie Database. Last modified 2015. Accessed April 17, 2025. <https://www.imdb.com/list/ls000097399/>.
- Seattle Daily Times (Seattle, Washington), April 12, 1961. America's Historical Newspapers.
- Seidenbaum, Art. "Afterthoughts on Apollo: The Moon Vs. Mankind?" Los Angeles Times (1923-1995) (Los Angeles, Calif.), January 3, 1969, 1-c9. ProQuest Historical Newspapers: Los Angeles Times (156103802).
- Sherrard, Alfred, Stuart A. Bernstein, Ewa Jaxa-Debicka, Sargeni Child, Marilyn C. Trainer, Mildred B. Lerman, Robert A. Mullen, and Roger B. Douless. "Letters to the Editor: Fallout on Snow." The Washington Post, Times Herald (1959-1973) (Washington, D.C.), April 23, 1961, 1. ProQuest Historical Newspapers: The Washington Post (141304458).
- Shreve, Bradley G. "THE US, THE USSR, AND SPACE EXPLORATION, 1957-1963." International Journal on World Peace 20, no. 2 (2003): 67-83. JSTOR.
- Siddiqi, Asif. "The Rockets' Red Glare: Technology, Conflict, and Terror in the Soviet Union." Technology and Culture 44, no. 3 (2003): 470-501. JSTOR.
- Siracusa, Joseph. "Bay of Pigs." In America in the World, 1776 to the Present: A Supplement to the Dictionary of American History, edited by Edward J. Blum, 125-26. Vol. 1. Farmington Hills, MI: Charles Scribner's Sons, 2016. Gale eBooks.
- "Space Race: What Prompted the United States to Venture into Space?" In American Social and Political Movements, 1945-2000: Pursuit of Liberty, edited by Robert J. Allison, 241-48. Vol. 2 of History in Dispute. Detroit, MI: St. James Press, 2000. Digital file.
- Special to THE NEW YORK TIMES. "Peace and Arms Pledges of the United States Given to the World Disarmament Parley by Davis." New York Times (1923-) (New York, N.Y.), May 23, 1933, 1. ProQuest Historical Newspapers: The New York Times (100649182).
- The Times. "London Ordeal of V2s." April 27, 1945, 2. The Times Digital Archive.
- The Times of India (1861-) (Mumbai, India). "Yuri Gagarin in Lucknow: WARM WELCOME ACCORDED." December 2, 1961, 5. ProQuest Historical Newspapers: International Collection (613289663).
- Van, A. Bowdoin. "Project Apollo." In Encyclopedia of the Sixties: A Decade of Culture and Counterculture, edited by James S. Baugess and Abbe Allen DeBolt, 536-38. Vol. 2. Santa Barbara, CA: Greenwood, 2012. Gale eBooks.
- "Vietnam War." In Gale U.S. History Online Collection. Detroit, MI: Gale, 2024. Gale in Context: U.S. History.
- Vohra, H. R. "Wave of Self-Criticism Sweeps U.S.: DEFEAT in SPACE RACE ACCEPTED SPORTINGLY the Times of India News Service." The Times of India (1861-) (Mumbai, India), April 15, 1961, 7. ProQuest Historical Newspapers: International Collection (741097163).
- Werth, Karsten. "A Surrogate for War—The U.S. Space Program in the 1960s." Amerikastudien / American Studies 49, no. 4 (2004): 563-87. JSTOR.

A Cross-Cultural Analysis - Effect That Advertising Has on Socialization Focusing on Skin Tone and the Perception of Women Aanya Nadkarni By Aanya Nadkarni

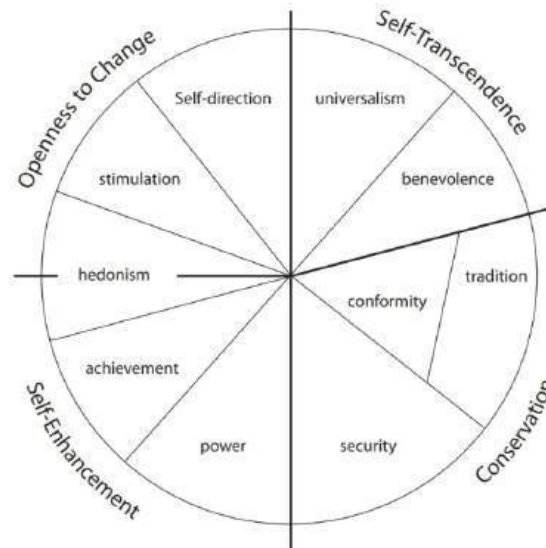
Abstract

Advertisements affect cultures' social norms significantly, but also have an effect on certain cultures' cultural values, and this can help to identify similarities and differences between separate cultures. Using the Schwartz Theory of Basic Values as a framework, this paper analyzes how hidden messages within advertisements reflect and portray values such as conformity, tradition, achievement, and self-direction, based on the messages hidden within skin lightening advertisements and portrayals of female models in advertisements. By performing a cross-cultural analysis with countries such as Japan, China, India, The Philippines, and The United States, the research emphasizes how skin lightening advertisements enforce colorism while associating lighter skin with success, beauty, and social mobility. Furthermore, this paper additionally studies the development of women's portrayal in advertisements over the years, focusing on sexualization, stereotyping, and the emergence of "femvertising". By comparing Western and Eastern advertising practices, the research finds that, while some beauty ideals are universal, cultural values have a significant role in shaping how women and skin tone are portrayed in advertisements. Altogether, this research highlights the role of advertisements as cultural agents, which both reflect and influence societal values, which contributes to the persistent beauty standards and gender norms across cultures.

Section 1: Background and Definitions

Socialization is the process by which individuals internalize social norms from the world around them. This means that, often, many of the unwritten rules society creates to dictate what behavior is considered acceptable are based on what is seen on social media. The media is filled with content and advertisements that businesses use to promote their products, and the average person is exposed to various advertisements a day. Advertisements and the way things are marketed are influenced by social norms, and these advertisements, in turn, shape the group's overall opinions, further ingraining these ideas in society.

The link between social norms and advertisements can be seen in many instances. For example, an increase in unhealthy food advertisements caused a spike in obesity, and an increase in tobacco and alcohol ads caused more underage drinking (Wilcox et al. 6). Furthermore, advertisements promote the qualities and functions of a product, and simultaneously convey "hidden messages" that diffuse cultural values and norms (Chen et al. 128). Advertisers use this feature to implant ideas in consumers' heads that can help increase sales. Ads for the skin-lightening industry often portray people with darker skin in negative ways, which further benefits the industry by fostering a desire among consumers for fairer skin.



Schwartz Theory of Basic Values
(Schwartz)

Advertisements significantly affect cultures' social norms and can also shape certain cultures' values, helping identify similarities and differences between cultures. The Schwartz Theory of Basic Values outlines the ten universal values for all humans: self-direction, stimulation, hedonism, achievement, power, security, conformity, tradition, benevolence, and universalism (Schwartz, pp. 5-7). These values are relevant in each different country's culture, and can be found in the messages hidden in advertisements. They can do numerous things, including helping to identify the origins of specific social issues. However, in this research, they help us understand which cultural values the West and the East share, based on their opinions on Women and Darker skin, which can be conveyed through advertisements. A good example of this is that the Philippines, a culture greatly influenced by the West, has values focused on change and self-enhancement, while Taiwan, a culture greatly influenced by Confucian and Japanese culture, values conformity and self-transcendence (Chen et al. 134).

A topic worth discussing is the difference between cultural values, social/cultural norms, and, subsequently, cultural practices. Cultural values are mainly the "inner person," or the ideas a person follows. At the same time, social norms are the "outer person," or the person one conforms to based on what is considered acceptable (Frese 1328). The leading cause could be that cultural values tend to be abstract ideas, such as those outlined in The Schwartz Theory of Basic Values. Social norms are "rules" that society creates and can compel people to participate in cultural practices. Cultural practices are actions people engage in to conform to social norms (Frese 1328). An example of this is when children are told to brush their teeth twice a day, something told to them by others and, hence, a social norm. They comply, and eventually they get to a point where they even feel uncomfortable when they cannot brush twice, developing a routine, which is an effect of cultural practices (Frese 1328).

Another difference between norms and values is that values focus more on what actions are morally good or bad. Norms are unwritten rules made by society, which the average person follows in their daily life, while values are intangible or abstract ideas that people base or almost “judge” their lives on (Frese 1328). Furthermore, values are distinct from behavior, unlike norms (Frese 1328). A person may have a short-term goal, such as getting a promotion, which is an example of social norms. Contrastingly, values are goals that continue throughout a person's life, for example, vegetarianism, to fulfill the traditions of their religion. Based on this, the importance of cultural values in the context of this research can be expressed, and we can understand that it is important to consider both cultural norms and values when researching, given the different aspects of each in relation to society.

Section 2: Overview

The effect of advertisements on the desire for fairer skin

Whether during the Meiji era of Japan where lead-based foundation was used to make skin appear white (Li et al. 445), or in ancient India where a paste made of various roots and seeds was mixed and dried between two cakes of buffalo dung to “make the skin a shade or two lighter” (Patkar 135), there have always been ways to fulfill the desire for fairer skin. The question is: why is the concept of skin lightening so prominent in Asian countries today?

There could be a few reasons for this stigma, but this paper focuses specifically on how advertisements have contributed to it. One of the main reasons that there is such a controversy on skin tone could be because of how frequently Caucasian models are used in advertisements in Asia. Oftentimes, the models in advertisements for Asian beauty or skincare products are caucasian or have caucasian features, which can distort the idea of “beauty” for the audiences of these advertisements. In South Korea and Japan, the amount of caucasian models used in global advertisements hovers between 40-50% (Li et al. 447), and in Singapore, the amount of caucasian models used in advertisements is 73.3% (Frith et al. 57), which further establishes the influence that Western culture has on the East.

Additionally, there has always been a negative connotation to having dark skin in countries all over the world, not just Asia. In Indian culture, and many other cultures, black is associated with being “underprivileged” and “dirty”, while white is associated with positive messages (Li et al. 445), and these ideas are spread and promoted through the use of advertisements. An advertisement for Pears soap depicted an African American child being “cleaned” by Pears soap and turning into a white child, which expresses the link between darker skin and a perception of dirtiness. Furthermore, an ad by Emma Watson for Lancôme in 2013, which promoted a skin-lightening product called “Blanc Expert,” demonstrates the globalization of Western beauty standards to an Asian audience specifically. This is because the target market for this ad was Asian people, and even though the product was meant to be used to get rid of “dark spots,” it promotes the idea that anything dark is bad. Products like this were used not only to get rid of dark spots but also to lighten the skin overall.

There is also the idea that having pale, or whiter, skin is associated with success, a status people with darker skin would otherwise not have. An advertisement done by Seoul Secret in 2016 for the Seoul Secret Snowz Pill, a skin lightening pill, had actress Cris Horwang warn her audience that if she had stopped taking care of herself “the whiteness [she] had invested in may be lost.” Then, her skin began to darken into black, and a younger, fair-skinned counterpart appeared beside her. The ad ended with the slogan, “You just need to be white to win.” This advertisement not only promotes the idea that you need fairer skin to be successful, but also suggests that fairer skin is more desirable and makes a person look more youthful.

Many other advertisements promote ideas similar to this, such as the Watsons (2017) ad that expressed the idea that lighter skin was necessary to gain a man's affection. This Malaysian ad was a 15-minute-long video called “Lagenda Cun Raya,” based on the legend of Dayang Senandong, a princess with a beautiful singing voice who was cursed to have dark skin. In the advertisement, an aristocrat is looking for a potential consort and is enchanted by Dayang's beautiful singing voice, but is appalled to learn that she has dark skin. Dayang then washes her face with a Watson product, and her skin turns fairer, prompting the aristocrat to sigh in relief and agree to marry her (MalayMail). In the original folklore, there is greater emphasis on the concepts of “true love” and “inner beauty,” where the king falls in love with her for her beautiful singing voice and personality (MalayMail). The curse that darkens her skin is only lifted after Dayang gives birth to the king's child, sealing their love. This advertisement fosters the idea that women with darker skin are less likely to win the affection of a man, further solidifying the societal preference for fairer skin, and associates darker skin with ugliness.

Interestingly, advertisements for the Seoul Secret Snowz Pill and the ad “Lagenda Kun Raya” both convey the values of achievement, and go against universalism. The Seoul Secret Snows Pill ad depicts fairer skin as a way for people to move up in society, and “Lagenda Kun Raya” shows that having fairer skin can help someone attain a better love life. Because of this it can be seen that both ads go against the value of universalism since they are indirectly portraying having darker skin as a disadvantage and giving privilege to some people, while demeaning others.

The cultural values present in all of the skin lightning advertisements aforementioned are mainly conformity and tradition. The advertisements are stressing that skin lightning is a necessary evil to be able to maintain the beauty standard that society has traditionally believed to be beautiful for centuries, condemning women who have darker skin and ensuring that there is a very negative association towards people with darker skin. This creates a harmful power dynamic between people with fairer and darker skin, and can cause people with darker skin to be looked down upon by society, restricting their opportunities.

Moreover, due to the prevalence of skin lightening in East Asian countries, there are whole companies that base their brand on the notion that having fairer skin is “superior”. One company like this is Glow & Lovely, which, until its 2020 rebrand, was known as Fair & Lovely. This company owns 70% of the market share in India's Skin-Lightening industry and has been known to use harmful stereotypes in its product promotions (Gajanan). Glow & Lovely used

rebranding tactics to maintain its relevance and combat colorism by changing its name, but it still sells “intense whitening” face wash and uses fairer models in its advertisements (The Seattle Medium). Unilever, the parent company of Glow & Lovely, even decided to remove the word “fair” from all of their products, but despite this, the harmful ideas promoted by their products continue to spread, and colorism persists (Gajanan)

Furthermore, the rebranding of many companies mainly applies to the West, not to Asian countries. Besides Unilever, companies such as L’Oréal and Nivea have used similar rebranding tactics to Glow & Lovely, either distancing themselves from the words “whitening” and “fair” or removing those terms from their products (The Seattle Medium). However, these claims only seem to apply in the West, specifically in the US and Europe. In Asian countries such as China, India, and Japan, L’Oréal uses many terms that even intentionally associate white with beauty, such as the term “Bihaku”, a term used in Japan that is used to describe “beautifully white” skin care (The Seattle Medium). Additionally, Nivea’s websites that were regional to Asia all contained extensive FAQs that recognized the connection to Asian beauty and skin lightening, and said that the company did not promote skin lightening at all, despite the numerous products that were marketed in India as “whitening” and the entire skin lightening section in the Malaysia website that used fairer models (The Seattle Medium). All of these promotions of skin lightening are not even remotely prevalent in western society, which further explains the differences in beauty preferences between the east and the west, and also says something about the beauty norms in the West. Advertisements are less focused on skin tone and many companies do not promote skin lightening as avidly in the West compared to the East, confirming that western beauty standards have a lesser focus on skin tone.

Cross-cultural analysis

Despite the general appeal for fairer skin in Asian cultures, many cultures have different reasons for lightening skin. For example, the use of lightening cosmetics in Japan is influenced by Western society but not used in the same way as in the West, where the goal is to enhance beauty (Li et al. 445). Instead, lightning products are used in ways that help people fit in with society's beauty norms (Li et al. 445). More than 90% of middle-class Japanese women whiten their skin to fit the standard, and even if they tan their bodies, they avoid their faces and continue to use the standard shade of foundation, which shows that skin lightening has become part of Japanese racial identity (Li et al. 445). From this, it can be seen that some of Japan's central cultural values are conformity and tradition, as most people lighten their skin to align with Japan's cultural traditions and identities. This view is similar to that of China and subsequently Taiwan. In China, lighter skin is seen as a sign of luxury, success, and prestige, and also as a sign of bodily control (Chen et al. 134). The desire for lighter skin has been prevalent in China for centuries, and these ideas have been passed down from generation to generation and expressed in advertisements, which help identify China’s cultural values of tradition, conformity, and achievement (Chen et al. 134). Subsequently, the cultural values of Taiwan, a society influenced mainly by Confucianism, are similar to those of China (Chen et al., pp. 133-135). Contrastingly,

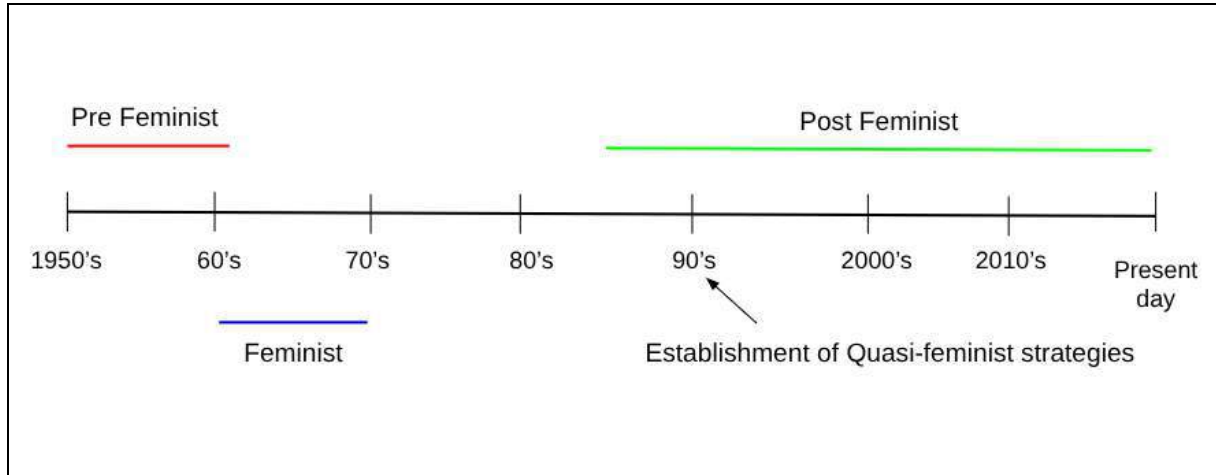
the Philippines has quite different cultural values from those of China, Taiwan, and Japan, even though skin lightening remains an essential aspect of societal standards of beauty (Chen et al. 134). The difference in the cultural values of the Philippines (Achievement and Hedonism) and those of Taiwan, Japan, and China (Tradition and Conformity) may be because the Philippines is influenced much more by Western society than the latter countries (Chen et al. 134).

Something that separates India from Western society is the prevalence of Bollywood (Li et al. 447). It was previously mentioned that models are often Caucasian in countries such as South Korea and China, but this is not the case in India, where popular Bollywood actors and actresses usually model for advertisements. Because of this, Indian beauty standards are less influenced by Western standards or Confucian standards like those in the countries mentioned above. Despite this, the cultural values of India are quite similar to those of China and Japan, since similar ideas are expressed in Indian skin lightening advertisements as compared to Chinese or Japanese skin lightening advertisements, but, one thing to note is that, unlike Japanese and Chinese people, Indian people do not lighten their skin to conform to a cultural identity, they lighten their skin for the opposite reason, to stand out. This is because it's generally less common for people of Indian origin to have fairer skin. From century to century, the desire for fairer skin increased and it became an idea that was passed down from generation to generation. Additionally, the want to stand out and be different from the norm shows that Indian people value novelty, which motivates them to move forward. This specifically shows an underlying value of self-direction which comes with this value. Based on this, it can be determined that India's cultural values are similar to Confucian ideals but are a blend of Western and Eastern ideals. The central cultural values of India are tradition, achievement, and stimulation.

The Effect of Advertisements on the Perception of Women

When the average person thinks of a stereotypical woman, they imagine a submissive, obedient housewife whose focus is to become a good mother and tend to her family. This perception is both widely incorrect and misguided. The reason for this traditionalist view, which many people still hold, is the media people were and are still exposed to. Advertisements significantly shape gender roles while also showing both the reality of the current moment and a glorified picture of the world (Sharma 236). The way that women are portrayed in advertisements directly influences the average person's perception of women and their roles in society. Because of this, studying advertisements from different time periods can help us understand how opinions about women in society change over time. For example, it would be much more common to see an advertisement portraying a woman as a housewife, say, in the 50s, than in the present day.

Development of Women in Advertisements



A timeline of different eras of women's portrayal in advertisements

Historically, women in advertisements have been stereotyped for almost two centuries (Sharma 237). Back in the 1980s, women were still portrayed as stereotypical homemakers (Soni 20). Further back, in the 1950s, women were portrayed similarly, but also as very weak and of lesser intelligence. A good example of this is the 1953 Alcoa Aluminum ad that promoted their Hytop bottle cap. The advertisement portrayed a woman holding a ketchup bottle, gently unscrewing its cap with her finger and wearing a surprised, delighted look. The advertisement's slogan was "You mean a woman can open it?" The ad did this to show how easy it was to unscrew the bottle cap by joking that even a woman could open it, which it was.

The 1950's to the early 1960's was considered the pre-feminist era of advertising, a time when most advertisements expressed the idea that women were meant to stay at home, they were not meant to make important decisions or do important things, were dependent on men, and even that men only saw women as sex objects and nothing more (Mager and Helgson 239). Concerns on these ideas by feminists led America into the Feminist Period of advertising, a time that lasted from the 1960's-70's (Mager and Helgson 239). This period in time challenged the notion of their being "masculine" careers, and advocated for more sexual freedom, paving the way for the postfeminist period. The postfeminist period has been ongoing ever since 1985, and has reflected some of the notions of the feminist period. While the advertising industry had come a long way since the pre-feminist era, stereotypes persisted in the media, and corporations were beginning to develop new strategies to appeal to both feminists and traditionalists (Mager and Helgson 240).

The strategy developed was a quasi-feminist strategy called "Femvertising," which has been developed and prominently used in industries targeted towards women, such as makeup products and menstrual products, to maintain relevance and appeal to customers. "Femvertising," or "female empowerment advertising," is an advertising strategy that portrays women in empowering positions to increase appeal to women (Sharma 237). This strategy is beneficial for reducing stereotypical portrayals of women, since the whole point is to empower women, but oftentimes "femvertising" is used by companies as a front to appeal to customers and is not fueled by actual support for women.

Sexualization of women

It is common for women to be sexualized in advertisements despite the nonsexual nature of the product being sold (Soni 20). A good example of this happening in real life is the recent campaign that American Eagle Jeans had in collaboration with Sydney Sweeney. The campaign was based on the 1980 Calvin Klein Jeans ad which featured Brooke Shields putting on a pair of jeans while talking to the camera. In the Sydney Sweeney version she is less exposed, but still doing the same, sexualised action. In the ad Sweeney is talking about genetics and says, "Genes are passed down from parents to offspring, often determining traits like hair color, personality and even eye color. My jeans are blue." The advertisement then ended with the slogan "Sydney Sweeney has great jeans". The controversy from this ad comes not only from the sexualised nature, but also because of the content of the ad. The ad that this was based on, with Brooke Shields, talks about breeding out "undesirable" genes, and uses Shields as a benchmark for what is desirable (Restrepo et al.). Because of this, many people thought that this ad was not only sexualizing women, but also saying that the desirable woman was caucasian, had blue eyes, and blonde hair.

The appearance of women changes whether or not they will be featured in an advertisement. Most of the time women chosen for advertisements will fit or be impossibly above the beauty standard (Soni pp. 20-21). Often, portrayals of women in advertisements will be so outlandish that they can distort women's ideas about body image. A woman's age has a very large influence on whether she will be featured in an advertisement. It is common for the women featured in advertisements to be from the ages 18-35 (Soni 26). It is very rare that you'll see a woman around the age of 60 in an advertisement, but it is more common to see a man of that age in an advertisement (Soni 21). This shows that women in advertisements are inaccurate to the average woman, and the models are focused more on appealing to the male gaze, than promoting the product ethically and efficiently.

In advertising, there have been around eight categories created to distinguish the portrayals of women in advertisements and the associations each advertisement seeks to make with the product. For this "mood" to be portrayed, advertisers take advantage of women's appearance and in what ways they are attractive (Solomon et al. 23). For example, when a person is described as "cute," their beauty is very different from someone who is described as "sexy". In a study done by Solomon, Michael R, Ashmore, Richard D, and Longo, Laura C, where magazine editors sorted different advertisements into different categories of beauty, the categories identified were Classic Beauty, Cute, Sex-Kitten, Sensual, Girl-Next-door, exotic, Feminine, and trendy (Solomon et al. pp. 27-28). There are some similarities between these categories, for example, the "Sex-Kitten" category and "Sensual" category both have the same sexual appeal, but "Sex-Kitten" refers to a less mature look. Likewise, the "Cute" and "Girl-Next-Door" categories share a similar atmosphere, making the models look more plain and focusing more on their emotions. The "Feminine" category is similar to the "Classic Beauty" category, which focuses on a "perfect" look. It is the opposite of the "trendy" category, which allows imperfections to resonate with the audience.

Using these categories, we can investigate how women are portrayed across cultures and what this reveals about each country's cultural values. Additionally, the ways Western women are viewed in Eastern cultures and vice versa can be examined based on the categories in which these women appear most often in advertisements.

In a study done by Katherine Toland Frith, Hong Cheng, and Ping Shaw, it was found that, when studying magazine advertisements of women from the most popular magazines in China, Singapore, and the U.S. that, while both Western and Asian cultures had around a 50% margin of advertisements in the “Classic Beauty” category, Asian cultures had a higher majority of ads that fit under the “Cute/Girl-Next-Door” category, while western cultures had a greater majority of “Sensual/Sexy” advertisements (Frith et al. pp. 57-58)

Interestingly, models in both Western and Asian cultures were found to be more often dressed in demure clothes, compared to erotic clothes; however, Asian models were posed more frequently in demure garments (84.5% margin) than western models (76.8% margin). In comparison, western models were posed more frequently in erotic clothes (23.7% margin) compared to Asian models (15.5% margin) (Frith et al. pp. 57-58).

Additionally, the study investigated whether the representation of the models in these advertisements was related to the product types being advertised. It was found that Western models appeared more frequently in clothing advertisements, while Asian models appeared more often in beauty product advertisements (skin and hair care). This could be because Western beauty standards are more focused on the body, and Asian beauty standards have more of a focus on facial features and skin tone, which explains why the data was skewed this way.

Something to consider with this study is that the categories “Western” and “Asian” do not necessarily mean that the advertisements in the “Asian” category all come from Asian magazines, and similarly for the “Western” category. The categories separate advertisements featuring Western and Asian models, allowing this study to analyze perceptions of Western women in Asian countries and vice versa.

Analysis

It was found in the study that both Western and Asian models were portrayed often in the “classic beauty” category, which shows that there are universal types of beauty that advertisers use between both Western and Asian cultures (Frith et al. 58). However, it is more often that Western models will be more sexualized compared to Asian models. When advertisers are promoting products that have sexual connotations, they will use Western models instead of Asian models, and western models will especially be used more often when the advertisement is targeted towards an Asian audience (Frith et al. 58).

Furthermore, the study found that Western models were more often featured in clothing advertisements, while Asian models were more often featured in hair and skin care advertisements (Frith et al. 59). This could explain the difference in beauty standards between Asian and Western cultures. Western beauty standards are more focused on the body, while Asian

beauty standards are more focused on facial features and skin tone, which falls in line with the fact that Western models are more sexualized.

Most advertisements featuring women are designed to appeal to the male gaze (Mager and Helgeson 240), so the fact that a large portion of models in advertisements from Eastern countries are portrayed in a more simplistic, “cute” style says something about how women are perceived in Eastern cultures. In the East, women are expected to be more conservative and less bold (Frith et al. 58). They are not generally expected to be in positions of high power in the workplace, and are almost diminished in their intelligence levels because of this “Girl-next-door” act. This explains why they are not generally put in erotic ads, when instead it is more common for Western models to be used in ads where the models are put in a sexual context, even if the advertisement is from an Asian country. This could be because in the East, it seems more acceptable for Western women to be sexualized, while Asian women generally are not, which shows restrictions in Asian women's sexual freedom (Frith et al. 58).

Comparatively, it is more common for advertisements to depict female models in sexual contexts in the West, an idea that can be inferred from the high proportion of advertisements that fell under the “Sensual/Sexy” category. While it is obvious that the values that would be conveyed from a high percentage of inappropriate depictions of women in advertisements, compared to the East, would be tradition and conformity, one thing that contrasts Western and Eastern cultures is that, while both try to diminish the standing of women by portraying them in a demeaning light, Western cultures choose to sexualize their models. In contrast, Eastern cultures choose to infantilize their models. This sexualization, despite how degrading it is, shows an openness for women to be more bold and expressive, and gives women more sexual freedom, something frowned upon in the East. Moreover, the margin of advertisements where Western women are portrayed in a sexual context is far less than half. Instead, it is more common for women across both cultures to be dressed in more demure clothes, which shows that Western advertisers do have a habit of sexualizing women, but still often gravitate to portraying them respectfully (Frith et al. 58). Because of this, the Western values that can be conveyed are achievement and self-direction.

Something worth noting is that sociologist Irving Goffman described different gazes in which many models often directed either towards or away from the camera (Frith et al. 54). For example, there is “licensed withdrawal”, where the model's gaze drifts away from the camera, and the “engaging gaze”, where the model looks at the camera with a sexually seductive gaze (Frith et al. 54). Interestingly, Samovar, Porter, & Jain, 2001 found that in Asian cultures, gazing towards the camera was seen as unfeminine and aggressive.

Based on these results, the cultural values of the West and the East can be determined. The wariness of the East to portray women in positions of power or in a sexual context highlights the East's conservative and traditional values. The habit of advertisers to portray Asian models as “cute” and “the girl-next-door” shows that Eastern cultures value conformity and tradition, as seen in the more modest atmosphere in which these models are portrayed in advertisements in these categories.

Conversely, the habit of advertisers in the West to portray their models in a bold, explicit manner highlights the West's openness to expression and to breaking boundaries. Opposite to conformity and tradition, the West values self-direction and achievement, which can be conveyed from the way that Western models are more often depicted in a way that gives them more autonomy over their sexuality and how much they are allowed to express it, despite the fact that western advertisements sexualize women as a way to almost objectify them.

Section 3: Research limitations & Future research ideas

An important topic to discuss is the limitations of the research. These limitations can skew the results of the research, but they can also present relevant ideas to study regarding this topic.

There is, for one, a lack of much research on the reason for the preference of tanner skin in western cultures, but instead more often comparisons of Western beauty standards to that of Asian cultures as a way to figure out the influence that western culture has on Asian societies (Li et al. pp. 444-449; Chen et al. pp. 127-158). Additionally, while there is much research on women's perceptions of advertisements (Soni pp. 20-29; Frith et al. pp. 53-61), there is little that analyzes the development of women's roles in advertisements over time, and even less that extends to the present day. This can be especially detrimental to the data, since we now live in a time where social media has now become a whole new marketing strategy, which can make the conclusions of this research a bit out of date.

Something that had such a large effect on the results of this research is the globalization of Western beauty standards. As Western ideologies have started to spread around the world, many traditional beauty practices from Eastern cultures have been forgotten so as to adopt beauty standards from western cultures (Chen et al. 128). As said earlier, countries influenced by the West will have cultural values similar to those of the West; for example, the Philippines has cultural values of achievement and self-direction. The same can be said about the values of countries influenced by Confucianist ideals (Chen et al. 134). Because of this, to provide in-depth conclusions about the differences between the cultural values of the West and the East, the extent of Western or Confucian influence in a given country should be taken into account.

Something that came up often when researching women in advertising was women almost being coddled into accepting women's stereotypical portrayal in ads. This could be the result of these ideas being cemented into their head from a young age (Wilcox et al. 5). Advertisers have started to aim their advertisements towards audiences of children, since they are the consumers most prone to advertisements (Wilcox et al. 5). They usually begin at ages 7 or 8. However, some even start at ages 4 or 5, since that is when a child first distinguishes an advertisement from other forms of media (Wilcox et al., pp. 5-6). This is important to consider when evaluating women's perceptions across different age groups. For example, people who grew up in the feminist era might have a less traditional view of women and their roles in society. However, someone who grew up in the pre-feminist era could have a more conservative and stereotypical standpoint on where women fit in society.

Section 4: Conclusion

Based on the conclusions of this research, we can discern the impact of advertisements on the values of specific cultures, learn about differences between Western and Eastern cultures, and deduce the importance of advertisements in shaping desires for fairer skin and perceptions of women. The stark contrast between the values of achievement, self-direction, and hedonism in the West, compared to the values of conformity, security, and tradition in the East, can help us put reasoning behind the differences in the social norms of these two cultures. Additionally, by studying advertisements and learning about the strategies of advertisers from different cultures, we can see how people from one culture view people from another. The importance of this research is that it can help identify ways to reduce the spread of the harmful stereotypes explored in this paper. From very young ages, people in Asian countries have been taught negative ideas about having fairer skin, and young girls from all around the world have been taught harmful stereotypes about women's roles and where women fit in society. Based on the conclusions of this research, we can now understand how social norms are transmitted through advertisements, so advertisers can design their ads to help dispel negative social norms, which could specifically help reduce negative connotations of darker skin and stereotypical portrayals of women.

Works Cited

- Chen, Chih-Ping, et al. "Mirror of beauty: Cultural values reflected in online skincare advertising in the Philippines and Taiwan." *Corporate Management Review*, vol. 37, no. 2, 2017, pp. 127-158.
- Frese, Micheal. "Cultural Practices, Norms, and Values." *Journal of Cross-Cultural Psychology*, vol. 46, no. 10, 2015, pp. 1327-1330.
- Frith, Katherine Toland, et al. "Race and Beauty: A Comparison of Asian and Western Models in Women's Magazine Advertisements." *Sex Roles*, vol. 50, no. 1-2, 2004, pp. 53-61.
- Gajanan, Mahita. "Unilever Fair & Lovely Name Change and Colorism in Beauty." *Time Magazine*, 26 June 2020, <https://time.com/5860313/unilever-fair-and-lovely-name-change-colorism/>. Accessed 14 December 2025.
- Li, Eric P.H, et al. "Skin Lightening and Beauty in Four Asian Cultures." *Advances in Consumer Research*, vol. 35, 2008, pp. 444-449.
- Mager, John, and James G. Helgson. "Fifty Years of Advertising Images: Some Changing Perspectives on Role Portrayals Along with Enduring Consistencies." *Sex Roles*, vol. 64, 2011, pp. 238-252.
- MalayMail. "Under fire, Watsons says blackface Raya ad 'highlights inner beauty, true love.'" *Malay Mail*, 2017, https://www.malaymail.com/news/malaysia/2017/06/07/under-fire-watsons-says-blackface-raya-ad-highlights-inner-beauty-true-love/1394473#google_vignette. Accessed 7 11 2025.
- Patkar, Kunda B. "Herbal cosmetics in ancient India." *Indian J Plast Surg Supplement*, vol. 42, 2008, pp. 134-137.
- Restrepo, Manuela López, et al. "The ad campaign that launched a thousand critiques: Sydney Sweeney's jeans." *NPR*, 1 August 2025, <https://www.npr.org/2025/08/01/nx-sl-5487286/sydney-sweeney-american-eagle-explain-ed-why-controversy-racist-eugenics-trump-bathwater-ad-klein-statement>. Accessed 15 December 2025.
- Schwartz, Shalom H. "An Overview of the Schwartz Theory of Basic Values." *The Hebrew University of Jerusalem*, vol. 2, no. 1, 2012, p. 20.
- Schwartz, Shalom H. "Understanding values: Schwartz theory of basic values – Integration and Implementation Insights." *Integration and Implementation Insights*, 10 May 2022, <https://i2insights.org/2022/05/10/schwartz-theory-of-basic-values/>. Accessed 14 December 2025.
- The Seattle Medium. "After Rebranding In The West, Many Beauty Companies Are Still Offering To 'Whiten' Skin Elsewhere." *The Seattle Medium*, 2022, [https://seattlemedium.com/after-rebranding-in-the-west-many-beauty-companies-are-still-offering-to-whiten-skin-elsewhere/#:~:text=Johnson%20&%20Johnson%20announced%](https://seattlemedium.com/after-rebranding-in-the-west-many-beauty-companies-are-still-offering-to-whiten-skin-elsewhere/#:~:text=Johnson%20&%20Johnson%20announced%20)

20it%20would,of%20our%20diverse%20consumer%20base.%E2%80%9D. Accessed 7 11 2025.

Sharma, Sangeeta. "Role Portrayal of Women in Advertising: An Empirical Study." *Journal of International Women's Studies*, vol. 22, no. 9, 2021, pp. 236-255.

Solomon, Michael R., et al. "The beauty match-up hypothesis: Congruence between types of Beauty and Product Images in Advertising." *Journal of Advertising*, vol. 21, no. 4, 1992, pp. 23-34.

Soni, Pallavi. "The Portrayal of Women in Advertising." *International Journal of Engineering and Management Research*, vol. 10, no. 4, 2020, pp. 20-29.

Wilcox, Brian L., et al. "Report of the APA Task Force on Advertising and Children." *American Psychological Association*, vol. 1, no. 1, 2004, p. 65.

Efficacy of the BNT116 mRNA Vaccine for the Treatment of Non-Small Cell Lung Cancer

By Chloe Tam

Abstract

Non-small cell lung cancer (NSCLC) is a form of lung cancer that can be treated with immunotherapies, including checkpoint inhibitors, chimeric antigen receptor (CAR) T cells, and cancer vaccines. Currently, several cancer vaccines are undergoing clinical trials for the treatment of NSCLC, including the BNT116 mRNA vaccine from BioNTech. This paper explores the mechanism of action of BNT116 and reviews clinical trials testing its efficacy and safety for NSCLC treatment. Data from ongoing phase 2 trials suggests that, in combination with other therapeutics, the BNT116 vaccine has potential to become a new treatment method for NSCLC in the future.

Introduction

Cancer and Immunotherapy

Cancer is one of the leading causes of death worldwide (Altman et al.). Cancer is caused by changes in the DNA known as oncogenic mutations (Bardeesy et al). Oncogenic mutations are classified through a common set of defining traits known as hallmarks. The hallmarks of cancer include immunoevasion, or the capability to avoid detection by the immune system. Other hallmarks include self-sufficiency in growth signals, insensitivity to inhibitory growth signals, the ability to invade tissues, also known as metastasis, unlimited replication potential, angiogenesis, and evasion of apoptosis (Hanahan and Weinberg).

Lung cancers can be subtyped as non-small cell lung cancer (NSCLC), small cell lung cancer (SCLC), or one of several less prevalent subtypes. NSCLC is the most prevalent subtype of lung cancer, comprising 80%-85% of lung cancer cases (Siddiqui et al.). Moreover, NSCLC is difficult to diagnose at its early stages when it may be asymptomatic. For this reason, there is an unmet need for more effective early diagnostic methods. As a result, NSCLC is often diagnosed at later stages with lower survival rates. Once detected, cancers like NSCLC can be treated with various therapies including chemotherapy, surgery, irradiation, and immunotherapy (Lahiri et al.).

Immunotherapy assists the immune system in detecting and destroying cancer cells. Multiple forms of immunotherapy have been approved for use in humans, such as immune checkpoint inhibitors, chimeric antigen receptor (CAR)-T cells, and cancer vaccines. Immune checkpoint inhibitors are antibodies that bind to components of immune checkpoints, preventing activation of signals that inhibit T cells and driving an immune response against cancer cells (Massafra et al.).

Cancer vaccines that promote antitumor activity by presenting cancer-associated antigens to the immune system have been developed and approved for use in humans. These vaccines induce the immune system to recognize and attack tumors such as NSCLC. Cancer vaccines

stimulate antitumor immunity by inoculating the patient with a tumor-associated antigen (TAAs) that can be taken up by antigen presenting cells. Cancer vaccines deliver one or more TAAs to the patient. There are numerous types of vaccines that have been developed, such as cancer cell-based vaccines, dendritic cell (DC) based vaccines, protein or peptide-based vaccines, nucleic acid vaccines, and personalized neoantigen vaccines. (Chen et al.).

mRNA vaccines are one class of cancer vaccines currently being studied for the treatment of various cancers. Upon injection, mRNA vaccines release mRNA encoding TAAs into the cytoplasm of a cell. The mRNA is then translated into protein, loaded onto the surface of antigen presenting cells, and used to activate cells of the innate system through pattern recognition receptors (Chen et al.) This paper will explore the mechanism of action of a specific mRNA vaccine, the BNT116 mRNA vaccine, and summarize a clinical trial testing its efficacy and safety as a monotherapy and in combination with other immunotherapeutic drugs in adults with NSCLC.

BNT116 Method

Current FDA Approved Therapies

Cemiplimab

One way that cancer cells can become immunoevasive is by overexpressing the cell surface protein programmed death-ligand 1 (PD-L1), which inhibits the cytotoxic activity of T cells. PD-L1 interacts with the receptor PD-1 on T cells to inactivate T cells and allow tumor cells to escape detection by the immune system. This interaction is classified as an immune checkpoint (Lahiri et al.). Cemiplimab is an immune checkpoint inhibitor consisting of an antibody targeting PD-1. Cemiplimab binds to the PD-1 receptor to inhibit PD-L1 attachment and prevent the activation of programmed cell death pathways in T cells. As a result, cytotoxic T cells can then bypass cancer immunoevasion to attack and kill cancer cells. However, cemiplimab treatment may be associated with side effects such as fatigue, headache, musculoskeletal pain, joint pain, diarrhea, nausea, vomiting, decrease in appetite, weight loss, fever, cough, dyspnea, pruritus, and rash. Additionally, 15%-25% of patients who were treated with immunotherapy experienced side effects related to immune activation (“Cemiplimab”).

Docetaxel and Paclitaxel

Docetaxel (DTX) and paclitaxel (PTX) are semisynthetic taxane chemotherapeutics that target microtubule assembly. Both DTX and PTX disrupt tubulin polymerization and breakdown, which stabilizes growing microtubules and disrupts mitosis, leading to cell cycle arrest. This slows the proliferation rate of tumor cells, which prevents the tumors from growing any further (Hashemi et al.; Wakelee and Gubens).

It remains unclear which chemotherapeutics might best enhance the efficacy of cancer vaccines like BNT116. Therefore, ongoing clinical trials are currently testing different combinations of BNT116 and chemo or immunotherapeutics.

BNT116 mRNA Vaccine

The BNT116 vaccine is an mRNA-based cancer vaccine developed by BioNTech that is delivered by lipid nanoparticles, which are microscopic particles made of lipids. The BNT116 vaccine was designed to treat NSCLC, and includes 6 different mRNAs encoding the proteins CLDN6, KK-LC-1, MAGE-A3, MAGE-A4, MAGE-C1, and PRAME— tumor-associated antigens that are often overexpressed in NSCLC. Early stage clinical trials suggested that the BNT116 vaccine demonstrated antitumor efficacy as a monotherapy or in combination with the immunotherapeutic drugs cemiplimab, docetaxel, and possibly paclitaxel (Dziadziuszko et al.).

Clinical Trials of the BNT116 Vaccine

Currently, the BNT116 vaccine is undergoing phase 2 clinical trials to assess for efficacy in humans. . The first and most recent clinical trial of BNT116 was the open-label LuCa-MERIT-1 trial (NCT05142189). In the LuCa-MERIT-1 trial, patients with NSCLC were treated with the BNT116 vaccine itself or in combination with cemiplimab, docetaxel, or paclitaxel (Dziadziuszko et al.).

The LuCa-MERIT-1 trial was divided into 9 cohorts in a 3+3 design, where 3 patients were enrolled to receive the lowest dose of a drug. If no severe side effects, also known as dose-limiting toxicities (DLTs), occurred in the first cohort, then additional patients were enrolled to receive the next highest dose (Kurzrock et al.). Cohorts 1-4 and 7-9 are currently ongoing for a total of 2 years, while cohorts 5 and 6 are ongoing for only 1 year. Each cohort received the first dose of BNT116, which was administered at 60 µg of RNA starting on the first day of the first cycle (Cycle 1 Day 1 [C1, D1]). Each cycle lasted for approximately 21 days. In the first cohort, which tested whether the BNT116 vaccine can work as a form of monotherapy, 5 other doses were given to the patients once a week during Cycle 1 and Cycle 2 upon the first dose. In Cycle 3, doses were administered every 3 weeks, and 350 mg of cemiplimab could be added if the patients' tumor expressed PD-L1. Additionally, the patients in cohorts 2 and 4 also received doses of cemiplimab starting from C1 D1 (Figure 1).

In all cohorts, the patients received weekly doses of 90 µg of RNA for the first 7 weeks. They also received a dose every 3 weeks on the first day of each remaining cycle. In the third cohort, the patients were given a 75 mg DXT administered intravenously every three weeks on the second day of each cycle (BioNTech SE). Cohort 4 included 20 frail patients with a median age of 69.5 years and an Eastern Cooperative Oncology Group (ECOG) score from 0-2 who had advanced or metastatic NSCLC. These participants each received the BNT116 vaccine alongside cemiplimab, which was distributed every 3 weeks (Dziadziuszko et al.). Currently, only data from cohorts 3 and 4 have been reported (Öven et al.; Dziadziuszko et al.).

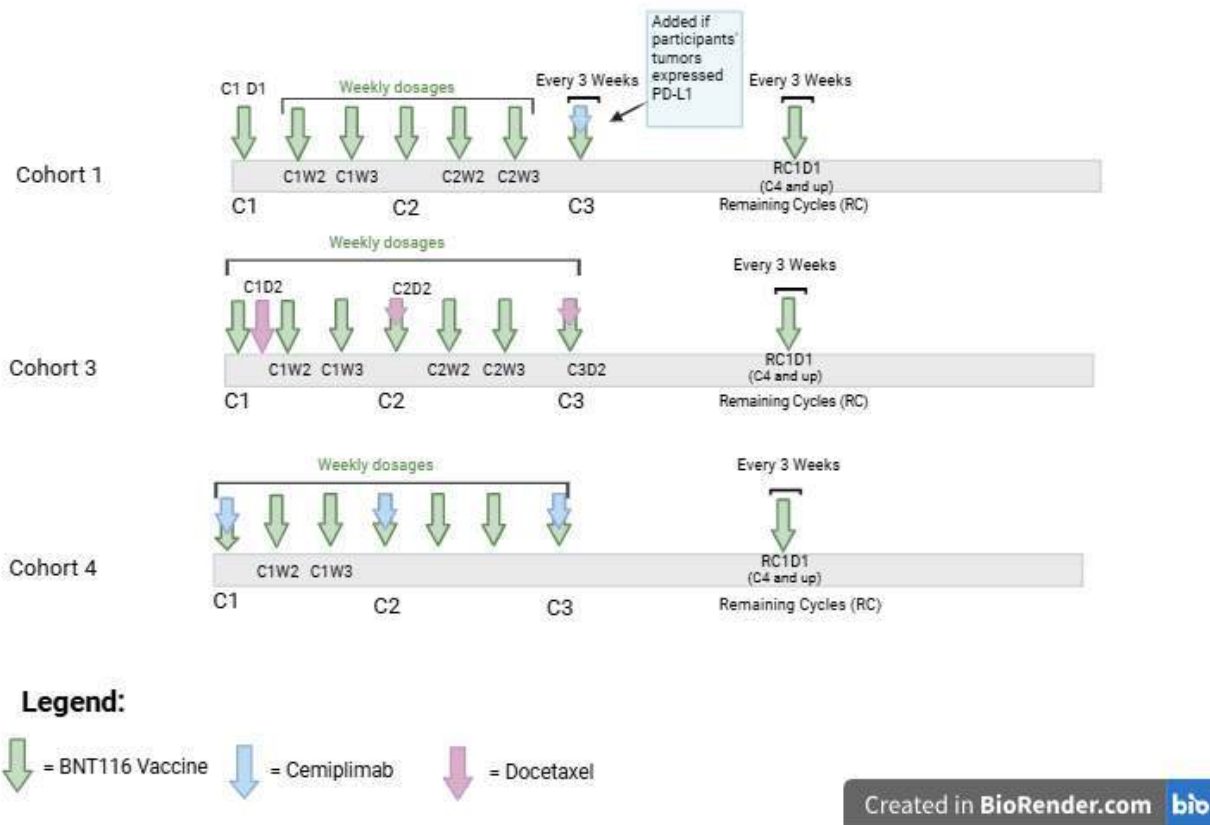


Figure 1. Different Cohorts and Drugs Administered for Participants in the LuCa-MERIT-1 Trial- The figure above depicts the type of treatment that participants receive in cohorts 1, 3, and 4. One dosage of the BNT116 vaccine is initially given to each cohort on C1 D1, then more doses are administered once a week for 7 weeks in total. In the third cohort, DTX is administered on the second day of each cycle, while doses of cemiplimab are administered every 3 weeks in Cohort 4. Created with <https://BioRender.com>

Results

Results of Cohorts 3 and 4

In the third cohort of the LuCa-MERIT-1 trial, all (100%) patients experienced a Treatment-Emergent Adverse Event (TEAE). 10/20 (50%) patients experienced serious TEAEs. These TEAEs included neutropenia, pneumonia, hypertension, decreased lymphocyte counts, diarrhea, and fatigue. No deaths or DLTs occurred during the dose confirmation period. 7/20 (35%) patients had a partial response to the BNT116 vaccine and DXT, while 10/20 (50%) patients had stable disease. Ex vivo experiments showed that T cells from patients treated with BNT116 were able to mount an immune response to TAAs. A partial response can be classified as a decrease in overall target lesions by 30% (Villaruz and Socinski). The results of cohort 3 support the potential efficacy of the BNT116 vaccine in combination with DTX (Öven et al.).

In the fourth cohort in the LuCa-MERIT-1 trial, 13/20 (65%) patients experienced mild to moderate treatment-related adverse events, while 7/20 (35%) patients experienced serious treatment-related adverse events (TRAEs). All patients experienced treatment related side effects from the BNT116 vaccine, and these TRAEs included supraventricular tachycardia, hypoxia, and hypertension. The best response was a partial response in 9/20 (45%) patients while 7/20 (35%) patients achieved stable disease (Dziadziuszko et al.). The results of cohort 4 support a favorable safety profile and potential efficacy of the BNT116 vaccine in combination with cemiplimab in patients with NSCLC (Dziadziuszko et al.). Together, the results from these two cohorts suggest that the BNT116 vaccine, when paired with cemiplimab or DTX, may represent a promising treatment for NSCLC in the future.

Discussion

Overall, the results of the LuCa-MERIT-1 trial suggest that the BNT116 mRNA vaccine could be effective in the treatment of NSCLC (Dziadziuszko et al.; Öven et al.). Results also show that the vaccine can work in combination with other therapies or treatment methods.

A limitation of this paper is that it only expands on the results of Cohorts 3 and 4 and does not cover the findings of Cohorts 1-2 and 5-9, as the data from these cohorts is not yet publicly available. Due to the lack of complete trial results, it remains unknown whether the BNT116 vaccine will demonstrate efficacy in combination with other treatment options. The small sample size of the LuCa-MERIT-1 trial is an important limitation that should be addressed with larger trials in the future. The current number of participants makes generalization to a broader population difficult. This limits the vaccine's versatility, which hinders the process of determining whether or not it has more benefits for people compared to the risks there are upon usage. If the vaccine is found to have risks rather than benefits, the FDA will reject it. With more participants, there is a high possibility that there will be individuals with bodies that react differently or in a unique manner to the treatment compared to others. A more diverse pool of patients also means that there could be a variety of different cancer mutations in the trial. With more diversity in a large sample size, scientists can encounter new adverse effects that weren't seen before and figure out whether or not the vaccine can work for everybody, which can reveal areas that require modification to make it more accessible for all NSCLC patients. In turn, more deaths can be prevented, since there would be more ways to stop the disease from further progression. To address the small sample issue ahead of time, researchers could gather a specific number of participants and then begin the trial once they have reached the desired amount.

Conclusion

Evidence from the LuCa-MERIT-1 trial suggests the promise of the BNT116 vaccine as a potential NSCLC treatment option. This implies that the vaccine can work with other NSCLC treatments. Future research should test the efficacy of the BNT116 vaccine in combination with other treatment methods. Additional trials for the vaccine should also be conducted to search for

more safety signals such as new side effects. To do this, researchers should identify cohorts of patients to participate in the LuCa-MERIT-1 trial and test other NSCLC treatment methods with the vaccine. Since some immunotherapies such as immune checkpoint inhibitors (like cemiplimab) were already tested in the trial, scientists could test to see if the BNT116 vaccine can work well with other cancer vaccines, including DNA vaccines, dendritic cell vaccines, or other mRNA vaccines. They could also test the vaccine with other immune checkpoint inhibitors like pembrolizumab and ipilimumab. One reason for this is because the BNT116 vaccine can work with cemiplimab. Since the vaccine can work with cemiplimab, there is a chance that it can work with other checkpoint inhibitors. Having more available therapies that can work with the vaccine could lead to an increase in treatment options for NSCLC patients in the future.

Works Cited

- Altman, Brian J., et al. "From Krebs to Clinic: Glutamine Metabolism to Cancer Therapy." *Nature Reviews Cancer*, vol. 16, no. 10, Oct. 2016, pp. 619–34. www.nature.com, <https://doi.org/10.1038/nrc.2016.71>.
- Bardeesy, Nabeel, et al. "Clonal Expansion and Attenuated Apoptosis in Wilms' Tumors Are Associated with P53 Gene Mutations." *Cancer Research*, vol. 55, no. 2, Jan. 1995, pp. 215–19.
- BioNTech SE. LuCa-MERIT-1: First-in-Human, Open Label, Phase I Dose Confirmation Trial Evaluating the Safety, Tolerability and Preliminary Efficacy of BNT116 Alone and in Combinations in Patients With Advanced Non-Small Cell Lung Cancer. Clinical trial registration no. NCT05142189, clinicaltrials.gov, 8 July 2025. clinicaltrials.gov, <https://clinicaltrials.gov/study/NCT05142189>.
- "Cemiplimab." *LiverTox: Clinical and Research Information on Drug-Induced Liver Injury*, National Institute of Diabetes and Digestive and Kidney Diseases, 2012. PubMed, <http://www.ncbi.nlm.nih.gov/books/NBK548288/>.
- Chen, Shaoyi, et al. "The Current Therapeutic Cancer Vaccines Landscape in Non-small Cell Lung Cancer." *International Journal of Cancer*, vol. 155, no. 11, Dec. 2024, pp. 1909–27. Crossref, <https://doi.org/10.1002/ijc.35088>.
- Dziadziuszko, Rafal, et al. "Abstract CT013: Phase I Trial Evaluating BNT116, a TAA-Encoding mRNA Vaccine, in Combination with Cemiplimab in Frail Patients with Advanced Non-Small Cell Lung Cancer (NSCLC)." *Cancer Research*, vol. 85, no. 8_Supplement_2, Apr. 2025, pp. CT013–CT013. Crossref, <https://doi.org/10.1158/1538-7445.am2025-ct013>.
- Hanahan, Douglas, and Robert A. Weinberg. "The Hallmarks of Cancer." *Cell*, vol. 100, no. 1, Jan. 2000, pp. 57–70. www.cell.com, [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9).
- Hashemi, Mehrdad, et al. "Paclitaxel and Docetaxel Resistance in Prostate Cancer: Molecular Mechanisms and Possible Therapeutic Strategies." *Biomedicine & Pharmacotherapy*, vol. 160, Apr. 2023, p. 114392. DOI.org (Crossref), <https://doi.org/10.1016/j.biopha.2023.114392>.
- Kurzrock, Razelle, et al. "Moving Beyond 3+3: The Future of Clinical Trial Design." *American Society of Clinical Oncology Educational Book*, no. 41, June 2021, pp. e133–44. DOI.org (Crossref), https://doi.org/10.1200/EDBK_319783.
- Lahiri, Aritraa, et al. "Lung Cancer Immunotherapy: Progress, Pitfalls, and Promises." *Molecular Cancer*, vol. 22, no. 1, Feb. 2023, p. 40. DOI.org (Crossref), <https://doi.org/10.1186/s12943-023-01740-y>.
- Massafra, Marco, et al. "Immunotherapeutic Advances for NSCLC." *Biologics: Targets and Therapy*, Volume 15, Oct. 2021, pp. 399–417. DOI.org (Crossref), <https://doi.org/10.2147/BTT.S295406>.
- Öven, Bala Başak, et al. "Abstract CT051: Preliminary Results from LuCa-MERIT-1, a First-in-Human Phase I Trial Evaluating the Fixed Antigen mRNA Vaccine BNT116 +

- Docetaxel in Patients with Advanced Non-Small Cell Lung Cancer.” *Cancer Research*, vol. 84, no. 7_Supplement, Apr. 2024, pp. CT051–CT051. DOI.org (Crossref), <https://doi.org/10.1158/1538-7445.AM2024-CT051>.
- Siddiqui, Faraz, et al. “Lung Cancer.” StatPearls, StatPearls Publishing, 2025. PubMed, <http://www.ncbi.nlm.nih.gov/books/NBK482357/>.
- Villaruz, Liza C., and Mark A. Socinski. “The Clinical Viewpoint: Definitions, Limitations of RECIST, Practical Considerations of Measurement.” *Clinical Cancer Research*, vol. 19, no. 10, May 2013, pp. 2629–36. DOI.org (Crossref), <https://doi.org/10.1158/1078-0432.CCR-12-2935>.
- Wakelee, Heather and Gubens. “Docetaxel in the Treatment of Non-Small Cell Lung Carcinoma: An Update and Analysis.” *Lung Cancer: Targets and Therapy*, June 2010, p. 63. DOI.org (Crossref), <https://doi.org/10.2147/LCTT.S6499>.

Gender and Higher Education: Exploring the Academic Journeys of Female Students in Bangladesh By Ahsun Abdul Goffar

1. Abstract

This paper analyzes the economic, socio-cultural and institutional limitations that characterize the experiences of female students in the National University sector of Bangladesh, a sector of the higher education system that is extensively ignored even though it enrolls millions of women. The study applies a mixed-method approach, involving survey data of 400 female students and 15 face-to-face interviews in six NU-based colleges. Results indicate that the most powerful elements that influence the possibility of students to access and receive higher education are financial strain and harassment. According to students, they have trouble with the costs of enrolment, long and unsafe travels, and most of the time they are harassed especially when they are commuting. Even though the social-cultural norms are not observed in the quantitative findings as direct predictors, the qualitative stories allow us to understand the hidden impact of these norms in the form of expectations of early marriage, the lack of mobility, and gendered restrictions of academic and extracurricular participation. The interviews also reveal the lack of even-handed institutional leadership in the NU colleges as certain students may face evident disparities in how they and their fellow students are treated by the faculty as well as their chances of leadership. Meanwhile, a minority characterizes more accommodating campuses, pointing to internal inequalities of the system. On the whole, the analysis demonstrates that gender inequality within NU colleges is caused by the combination of economic hardships, safety issues, and the strong patriarchal culture. It suggests specific changes such as financial assistance, secure transportation, campus policies that are gender sensitive, and mentorship to promote the development of Bangladesh in SDG 4 and SDG 5.

2. Introduction

Bangladesh has achieved substantial advancement in girls' education through its primary and secondary school programs during the past several decades. The transition from secondary education to tertiary education faces challenges because of traditional male-dominated social rules which control student movement and wedding schedules and influence their professional development choices. The National University operates through college affiliations which create both operational and teaching difficulties while the centralized Universities maintain their own campus-based facilities. Students need to travel extended routes from their residences to reach their colleges while their educational experiences at different institutions differ between urban and rural settings. The specific section addresses how female students face particular gender-related challenges which include unsafe travel conditions and family limitations and financial struggles and potential campus harassment. The research investigates diverse educational experiences of National University of Bangladesh female students through a complete mixed-methods research design. The research investigates multiple elements which affect student learning at Bangladesh universities to discover ongoing obstacles and new

possibilities for student empowerment in this essential educational system. The research investigates multiple elements which affect student learning to create recommendations that will establish equal support for female students in Bangladeshi higher education institutions. Research studies about female undergraduate students at Bangladesh's National University system remain scarce according to Alam et al. (2024) in their review of existing literature. Research about gender inequality in education exists but there are limited studies which focus on Bangladeshi women who study STEM subjects in patriarchal South Asian societies (Rahman, 2024). The academic literature lacks research about undergraduate women who study STEM subjects at Bangladeshi institutions despite worldwide recognition of gender disparities in these fields (Rahman, 2024). The absence of research about female students' academic and personal challenges and their classroom and social environment experiences creates a critical knowledge gap for developing inclusive educational spaces (Sivakumar et al., 2024) (Rahman, 2024). The National University operates through more than 200 colleges which serve millions of students who determine the academic paths of most Bangladeshi women seeking higher education. The identification of gender-based obstacles in higher education requires immediate action to fulfill Bangladesh's national goals for SDG 4 (Quality Education) and SDG 5 (Gender Equality). The research combines qualitative and quantitative data to establish evidence which will help develop specific solutions and funding plans and institutional changes to build an inclusive learning space for all students.

Research question

What are the key socio-cultural, economic, and institutional challenges that affect female students' access to higher education under the National University of Bangladesh?

Objectives

1. Identify major gender-based barriers.
2. Examine the relationship between socio-economic factors and academic satisfaction.
3. Explore students' lived experiences to complement quantitative findings.

3. Literature Review

Research studies about female undergraduate students at Bangladesh's National University system have not received sufficient attention according to Rahman (2024). Research studies about gender inequality in South Asian patriarchal societies including Bangladesh exist but they do not specifically investigate the challenges and opportunities women encounter in this particular setting (Rahman, 2024) (Sarder & Rani, 2025). The research investigates how academic demands and social norms and university support systems affect female students throughout this extensive university network (Rahman, 2024) (Wahid et al., 2025). The National University's college network across different locations creates special academic and teaching requirements which might create distinct learning experiences for female students than what

occurs in traditional university settings (Tewari, 2018). Research shows that female students must handle multiple responsibilities between their studies and personal life which creates obstacles for their academic success and mental health (Sivakumar et al., 2024). The lack of female mentors and gender-based discrimination at school creates additional difficulties for students to develop their self-identity and psychological well-being (Rahman, 2024). The educational journey of women in STEM fields becomes more challenging because of career anxiety and impostor syndrome which requires detailed research about their academic experiences (Rahman, 2024). These variables are explored through quantitative and qualitative data analysis to fully understand how female students experience their education in this specific academic context (Hasan & Akter, 2023) (Wahid et al., 2025). The study bridges the gap between previous knowledge through providing a comprehensive analysis of challenges and opportunities faced by female students in Bangladesh National University system (Siddiky, 2020) (Wahid et al., 2025). This basic information will be useful for the formulation of targeted interventions and policies for an equitable learning environment among women students and their enhanced academic performance (Amin et al., 2020) (Deblasio, 2023). The research results are pivotal as they respond to the continuous gender-related educational imbalances that still exist in male-dominated societies (Hossain et al., 2023). The research evaluates existing educational policies to discover their effectiveness and recommends modifications which will create an inclusive learning environment for female students in higher education (Hua & Zhao, 2024) (Rahman, 2024). The research uses social justice principles and capabilities theory to detect female graduate student inequalities in educational access and achievement and opportunity access (Ellegaard et al., 2024).

While some studies have touched on gender inequality and the educational hurdles women face in South Asia, there's a noticeable gap when it comes to research on female undergraduate students specifically within Bangladesh's National University system. Most of the existing literature addresses gender issues broadly or focuses on centralized universities. As a result, the distinct challenges faced by students in the National University's affiliated colleges have been largely overlooked. Even more worrying is our overall lack of understanding of the mechanisms whereby academic stress, peer expectations and support structures at these institutions are intertwined to affect well-being, identity and achievement in female students. Without research that blends these methods and delves into such topics, we lose sight of the complex hurdles at play — say, whether people are juggling competing responsibilities in their home and work lives or confronting gender bias or nervousness about careers. Undoubtedly there is a research gap to be filled in order to gain a comprehensive and evidence-based understanding of what female students feel, experience and are being exposed to in this vast —and unique— educational setting. It is crucial to investigate how such systemic inequalities as particularly family and society ones impede educational advancement and place barriers in the way of women among their professional careers in Bangladesh (Sarder & Rani, 2025).

4. Methodology

In this research, we used both quantitative and qualitative methods of data collection—a mixed method approach. We scrubbed the internet for primary and secondary sources to dive deep into our questions. The research was focused on the National University — more than 200 colleges that educate millions of undergraduates.

4.1 Quantitative Component

Semi-structured questionnaires were used to collect primary data by applying the Google Forms tool. The questionnaires were administered to the participants who were present on their college campuses, but when some individuals could not visit the campus, they were given a chance to fill out the questionnaires online. Overall, 400 female students, between the age of slightly less than 20 and more than 25, agreed to provide their ideas concerning their educational experience (Table 1). All these participants were enrolled at the National University by October of 2025, and they are first-year undergraduates through graduate students. Particularly, 19.5 percent were first-year students, 24.3 percent were second year students, 38.3 percent students were in their third year, 13.3 percent were in their final year, and 4.2 percent students were in graduate programs (Table 2). The case of the secondary data involved a detailed literature review of peer-reviewed journal articles, reports, edited books, policy documents, and other sources of interest. These secondary materials helped to enhance our knowledge of educational experiences and also helped to triangulate the primary information.

Table 1: Demography of survey participants

	Characteristics	N= 400	Percentage
Age	Below 20	58	14.5%
	20-22	146	36.5%
	23-25	183	45.5%
	Above 25	14	3.5%

Table 2: Survey Participants Affiliation

	Characteristics	N = 400	Percentage
Student Affiliation	Undergraduate 1 st year	78	19.5%
	2 nd year	98	24.5%
	3 rd year	153	38.3%
	4 th year	54	13.5%
	Master's	17	4.2%

4.2 Sample Selection Criteria

There were three main sampling strategies used in the study since the researcher was gathering both the survey and qualitative field data. In the case of the survey part, simplicity of random sampling was integrated with convenience sampling, due to the implementation with the help of Google Forms. The target population consisted solely of female students who could be found in colleges that are associated with the National University. The Google Form was shared by representatives of the class, WhatsApp groups, the help of the faculty members, and some random face-to-face outreach. The sampling criteria excluded male respondents and those who considered their sexual orientations other than that of a female. The eligibility requirements would include that the participants have to be female students studying in a National University College and between 20 and 25 years. In in-depth interviews (IDIs) one further limitation was applied: the subjects in the IDIs were limited to female students who encounter socio-cultural limitations and can express their experiences and coping strategies at the institution. Based on these rules, we were able to conduct 15 IDIs, which were all respondents chosen in the survey. All of the interviews were face-to-face. Strict ethical considerations were followed; the participants volunteered and gave their consent to recording. Fourteen of the participants agreed to be audio recorded and one participant declined to record.

4.3 Data Analysis

In the case of the quantitative data, descriptive statistical methods were used to explain educational experiences in the research field. This discussion gave the percentage of the respondents who had experienced economic barriers which negatively influenced their education and living conditions. Standard text analysis of the obtained data was used to enhance the discussion of qualitative data, as in-depth interviews (IDIs) were implemented. The main theme was open-coding approaches, each line of text was systematically reviewed to identify differences and similarities in concepts and themes. Following thematic analysis allowed a derivation of qualitative thematic insights which were structured, based on themes, out of the qualitative data.

4.4 Limitations

Just like any empirical inquiry, this research was prone to several limitations. To introduce a little bias in the affiliation of colleges, first, despite the fact that National University has a population of more than 200 colleges, only six institutions were sampled which represent both urban and rural environments hence giving a narrow but illustrative view of affiliated colleges. Second, the voluntary character of the survey provided a rather limited sample, which produced non-response among some of the students who refused to provide their educational experiences. Lastly, the limited time means constrained the extent of data collection, which limited the size of the sample participants, and, by extension, the applicability of the results.

5. Quantitative Findings

5.1. Descriptive Overview

This section is now a quantitative analysis of the data taken of 400 female students attending different colleges of the National University of Bangladesh.

Age Distribution: 45 percent of the respondents were between the ages of 23 and 25 years with 36.25% between the ages of 20-22 years implying that the sample comprised relatively young women in tertiary education.

Household Income: Only an approximation of 48 percent of students surveyed belonged to a household with an income of 30,001 to 50,000 Bangladeshi taka per month with only 6.75 percent of the respondents having an income lower than 15,000 takas per month. This distribution indicates that it is made up of predominantly middle-income families with enough funds to enable higher education.

Parental Education: Approximately 38 percent of the respondents reported that their parents had completed higher secondary education with 24.25 percent reporting that one or another of their parents had a bachelor's degree or higher. Only a 1-percent respondents stated that their parents did not receive any form of formal education, so the level of parental education could have a positive effect on the choice of female students to pursue higher education.

Enrollment Challenges: 40.65 percent of the respondents reported barriers they had faced throughout the registration process, highlighting the overall structural and procedural issues that the National University system continues to face.

Experience of Harassment: 35 percent of all respondents said they had experienced occasional harassment, twelve five percent said they had experienced frequent harassment and 44.5 percent said that they had experienced rare harassment. Under such circumstances, more than eighty-seven percent of the students experienced gender-based harassment in the course of their university life.

Socio cultural norms: 41 percent of the respondents confirmed that socio-cultural barriers hindered their engagement in group or academic activities, which underscores the persistent nature of patriarchal norms that limit female participation in higher learning institutions.

Overall Satisfaction: 46.25 percent of the respondents indicated that they were neutral about their overall experience and 39 percent indicated their satisfaction. The percentage of dissatisfaction expressed was only 6.75 percent, which shows that despite the realization that the challenges are present, most of the female students get the level of institutional support at a moderate level.

5.2. Chi-Square Tests (Association Between Variables)

Table 3: Chi-Square Tests

Association Tested	Chi ²	p-value	Interpretation
Income vs Enrollment Difficulty	22.75	0.000	Highly significant — students from lower-income families faced more enrollment challenges , indicating economic constraints as a major determinant of access.
Parental Education vs Harassment	46.00	0.000	Highly significant — harassment levels decrease with higher parental education , suggesting that parental awareness and socio-economic status help reduce vulnerability.

Socio-cultural Norms vs Satisfaction	3.80	0.434	Not significant — satisfaction levels did not vary strongly with socio-cultural constraints, possibly because institutional experiences (not cultural norms alone) shape satisfaction.

5.3. Regression Analysis (Determinants of Satisfaction)

Model:

$$\text{satisf\_num} = \beta_0 + \beta_1(\text{income\_num}) + \beta_2(\text{parentedu\_num}) + \beta_3(\text{harass\_num}) + \beta_4(\text{norm\_dummy}) + \varepsilon$$

Table 4: Regression Analysis

Variable	Coefficient	p-value	Interpretation
Income Level	+0.117	0.029	Significant positive relationship — higher family income increases satisfaction with university experience.
Parental Education	+0.009	0.739	Not significant — education of parents does not directly influence perceived satisfaction.
Harassment Frequency	-0.137	0.001	Strong negative impact — increased harassment reduces student satisfaction and overall well-being.
Socio-cultural Norms	+0.056	0.521	Not significant — external social restrictions do not significantly alter perceived satisfaction within campus life.

Model Fit: $R^2 = 0.045$, $F(4,395) = 4.64$, $p = 0.0011 \rightarrow$ While the numbers seem modest, this model holds statistical significance. It demonstrates that income and harassment are some of the factors that affect the satisfaction of students.

Going further, we are able to discover that the most predictive variables of educational satisfaction are economic stability and personal safety, namely harassment, among female students. This realization is directly related to the emphasis on economic variables and social-cultural challenges.

5.4. Correlation Matrix

	enroll~m	harass~m	norm_d~y	satisf~m	income~m	parent~m
enroll_num	1.0000					
harass_num	-0.1233 0.0136	1.0000				
norm_dummy	0.0435 0.3850	0.0370 0.4601	1.0000			
satisf_num	0.1214 0.0151	-0.1782 0.0003	0.0316 0.5279	1.0000		
income_num	0.1611 0.0012	-0.1579 0.0015	0.0726 0.1475	0.1361 0.0064	1.0000	
parentedu_~m	0.0586 0.2426	0.0701 0.1618	-0.1401 0.0050	-0.0151 0.7637	-0.1439 0.0039	1.0000

Issues and Harassment in Enrollment: $r = -0.12$, $p = 0.013$. This suggests that those students who had issues with enrolling were a little less prone to report harassment, probably due to overall lower participation. Now what about harassment and contentment? The relationship is clear: $r = 0.18$, $p = 0.0003$. The study shows that the more the students are harassed, the lower the level of their satisfaction- strong negative correlation. Next, we shall take income: $r = +0.14$, $p = 0.006$ shows a positive correlation. A higher level of income is likely to boost the levels of satisfaction. But how about the education of parents? $r = -0.0151$, $p = 0.76$ indicates weak and insignificant relationship.

5.5. Factor Analysis (Principal Component Analysis)

Upon digging further, we identified two important factors that come to explain 50 percent of the total variance.

Factor-1 that we may term as Institutional and Economic Challenges presents high loadings of such issues as difficulty in enrollment, harassment, and satisfaction. Factor-2, which is Socio-cultural Constraints, is consistent with gender sensitivity and norm -dummy. Such groupings support the notion that the issues encountered by the students are not single-stranded and this is in line with your framework that categorizes the barriers into institutional, economic, and socio-cultural boundaries.

5.6. Composite Gender Barrier Index

Now, we have developed a composite measure which draws in enrollment difficulty, harassment, and socio-cultural constraints. The results? An average of 1.64 (SD =0.41), which represents moderate gender-related barriers in general. In regressing the index with the socio-economic variables, income ($p = 0.637$) and parental education ($p = 0.665$) were

insignificant. Satisfaction is one of the only variables that have a slight negative trend ($p = -0.093$): when satisfaction drops, so do perceived gender barriers. This points to the fact that the female students are not always secure in the hands of the family with good income or high level of education yet there are still barriers that are too deeply entrenched in the system.

5.7. Visual Results Summary



The statistical results present empirical evidence to support the hypothesis that economic variables, especially the family income, can have a significant influence in determining accessibility and satisfaction as regards to education. Institutional culture and harassment are major impediments towards the realization of gender equality. Despite the continuity of the socio-cultural norms, their direct effect upon academic satisfaction has subsided. The female students studying at the National University system are just going through the gendered educational process whereby the structural and institutional barriers are the primary hindrances to their advancement, other than necessarily their background. The Stata results supported the expected effects of income and harassment on the level of satisfaction of the female students about higher education. Additional factor analysis revealed that there were two dominant categories of barriers, namely, institutional/economic and socio-cultural. As much as income is expected to boost satisfaction, the adverse impact of harassment is high. These results highlight the urgency of gender sensitive policies and safe campuses. Overall, the findings are perfectly compatible with the mixed methodology of the study in question, which serves to highlight that

the problem of gender inequalities in the National University of Bangladesh is ingrained in social constructs and institutional failure.

6. Qualitative Findings

Interviews clarified the day to day struggles of the female students who are involved in higher education in the National University system. In line with the quantitative findings, three major themes were evident, including economic obstacles, harassment and safety issues, and socio-cultural expectations, which all intersect to create an academic experience and future. First, there are the economic issues and affordability which attract immediate attention. Although the quantitative data proved that income is the important predictor of satisfaction ($b = + 0.117$, $p = 0.029$), the interview stories revealed the tangible sufferings behind the numbers. Students over and over again mentioned the accumulative pressure of tuition fees, transport and learning supplies, which made study tedious. As an illustration, Nabila talked about how the distance of over 20km between her house and her college gave her commuting challenges that had a negative impact on the attendance of classes. On the same note, Aklima shared that she faces ***"financial conditions, mental stress, attendance requirements, and various economic factors that affect academic progress."*** These testimonies highlight the fact that even the students whose families belong to middle-income groups face barriers to academic participation and success. The affordability of the National University is also a relevant issue, as the participants who can't cover transportation costs but support family members, like Neha explained that she struggles ***"to cover transportation costs"*** while also contributing to family expenses, show us that it is not an unchallenged issue despite relatively lower tuition rates at the National University.

Second, one should not overlook harassment and safety issues. The Stata regression statistically proved the existence of a significant counterproductive influence of harassment on satisfaction ($b = -0.137$, $p = 0.001$). This finding was supported by interview testimony. Harassment, especially on the way to school, remained an issue that affected the mental health of the students as well as their perceived security. Raida reported that she was harassed on her way to the college, when older men tried to have inappropriate physical contact. Riddhi expressed concerns involving the threat of insecurity during the commute, such as threatening behaviors by other passengers. Interestingly, some of the interviewees said that they felt safe on campus implying that the institutional programs are promoting gender sensitive campuses. However, there is still a psychological cost of external threats which forcefully and subtly affects the engagement in the academic activities. Sumaiya Alam narrated a case of undesired actions of male students, but she was responsive to the views of the institution suggesting that well-working institutional action can alleviate the negative impact of harassment as postulated by the quantitative model. Last but not the least, the socio-cultural expectations and gender roles became salient issues. Although quantitative data showed that cultural norms did not have any direct impact on satisfaction ($p = 0.521$), qualitative texts helped to understand how they lingered in the background. The pressure to get married, lack of movement, and the deeply rooted gender roles came up as obstacles. Neha

revealed that she felt pressure in the society to get married and stop her education; Aklima reported about the twofold pressures put on her by the family and society. Even those students who felt there were little explicit restraints like Nabila who did not see any gender preference in her family admitted that there was widespread social pressure concerning early marriage. In addition, implicit sex-imbalances in extra-curriculum activities, as well as faculty treatment were pointed out. Noting that male pupils tend to have a better access to sports and leadership in clubs, Saba showed that the issues with the institutions may lead to gender constraints in the availability and the visibility of the opportunity. Overcoming these challenges, some students had positive institutional experiences which were aligned with the survey findings where 46 % of students chose a neutral answer and 39 % of students reported to be satisfied. A well-wish faculty, family approval, and non-discriminating peers seemed to cushion outer misfortunes. As an example, Marzan claimed that there was no gender inequality in her campus and that there was fair treatment of male and female students, and Nafisa did share a positive experience free of discrimination by her peers and faculty. On the other hand, some of them, including Sumaiya Binte Alam, raised differences referring to the fact that male students were more equipped with high-qualifying faculty, and overall, they were treated more fairly, which means that equality is not a universally achieved situation on campuses.

7. Discussion

The combination of the mixed-method results shows the distinct congruence between the quantitative and qualitative elements. The regression equation recognizes income and harassment as the most important variables that define female students' satisfaction with higher education. At the same time, the interviews offer a subtle picture of realities on the daily basis behind such statistics. Students with more affluent backgrounds claim less financial concerns and more satisfaction, and low-income students mention tuition, books, and transportation expenses as a big barrier on numerous occasions. This trend is reminiscent of the situation in Nabila where she was travelling more than 20km to college, which has been adversely affecting her attendance. Similarly, harassment, especially in transit, became an enormous negative predictor in the regression ($b = 0.137$, $p = 0.001$) and a common theme in the qualitative data. The description of the experience of older men making inappropriate advances by Raida highlights the impact of the severe safety issues on the level of satisfaction associated with the schooling. Combined, these results prove that economic resources and personal security are determining factors of the education experiences of women in the National University system. The sharp contrast is seen when one looks at the socio-cultural norms. Culturally, there was no significant relationship between cultural restrictions and satisfaction ($p=0.521$). However, the participants outlined these norms as ubiquitous and emotionally strenuous. Both Neha and Aklima said that there was societal pressure to marry and drop [their] education, and that the conditions of their society do affect perseverance and motivation even though they may not be reflected in the data. This known discrepancy suggests that the action of cultural influences can be indirect, and the perception of students towards their institutionalization, and not towards their satisfaction levels.

As a result, a combined methodology will show that patriarchal rules lie in the depths of quantitative naivety. These results are quite close to the existing research in Bangladesh. Similarly, Chowdhury and Ahmed (2021) have also found financial constraints and patriarchal family pressures as two obstacles to the continued education of girls. The persistent existence of gender-based discrimination in higher institutions of learning was affirmed by Islam (2023). The current research builds upon these findings by proving, using regression and factor analysis, that harassment and financial demands are not isolated but occur in a more extensive structure. In line with the research by Khan and Alam (2024), it can be concluded that admission does not mean empowerment. The qualitative descriptions of disrupted quality of faculty and not being included in extracurricular activities point to the need to have better-quality and gender-sensitive institutions. Additionally, the constant leaky pipeline, which is highlighted by Hossain and Rahman (2025), is also extremely evident in the everyday lives of students, and, according to Hossain and Rahman, most of them feel that the opportunities to lead and be members of their clubs are limited.

8. Policy Recommendations

The facts show there is a dire necessity of the changes in the National University scheme. To begin with, the implementation of financial aid and transportation allowances would make it much more affordable to students whose families have low income.

Second, the gender-sensitive safety interventions, including the observed transportation pathways, anti-harassment committees, and anonymous reporting networks, would resolve the common problem of harassment.

Third, gender sensitivity and equal allocation of teaching resources should be trained to the faculty to avoid inequality, like the one noted by Sumaiya Binte Alam that the faculty is better qualified to teach male students.

Fourth, the establishment of on campus job opportunities would allow students to learn to make money as they study, thus developing skills and lessening reliance on families especially among the female students.

Lastly, there should be academic mentorship and psychosocial counseling to help students deal with societal demands when it comes to marriage and stigma. This assistance would be able to turn the observed resilience in interviews into a reality in the form of academic success. Regardless of the structural impediments, based on the qualitative data, female students demonstrate impressive strength and flexibility, and Marzan and Nafisa can be regarded as examples of women with high expectations and happiness, as their hope and satisfaction are results of their diligence and family support. It is an indication of a changing gender balance in the higher education environment of Bangladesh: women are no longer passive consumers of inequality but are quite literally maneuvering around a world of limited choices. This resilience is consistent with the current studies which suggest that the education of the women in Bangladesh is a crucial type of self-empowerment in a patriarchal society. As a result, although

systemic inequalities still exist, the will of such students is a good sign of significant steps on the way to SDG 4 (Quality Education) and SDG 5 (Gender Equality).

9. Conclusion

To sum up, this mixed-method research explains the complex interaction of socio-cultural, economic, and institutional determinants that influence the experiences of female students in the system of the National University in Bangladesh. The results emphasize the current struggles of these women in a patriarchal South Asian society especially in relation to getting higher education and the gender gap. Although gender inequality has been discussed in general terms, there is a clear deficiency of specific studies, which identify specific issues and opportunities of female undergraduates in the extensive system of National Universities, in particular, in the framework of qualitative research aimed at comprehending the real experience of women in the context of national differences. There exist economic constraints, which include low household earnings and high education costs thus reducing female students' academic satisfaction and performance in the National University-affiliated colleges considerably. In the case of safety issues and harassment, over 70 per cent of respondents are directly involved, which generates fear, restricts movement, and damages dignity. In addition to quantitative measures, socio-cultural economic pressure on marriage and gender roles also limits the educational participation of women. Institutional quality is highly unequal among affiliates of the National University with some providing strong supportive cultures, some endorsing discriminatory policies, and there is a pressing necessity to have a system where a standard and accountability exist. Increasing safety, affordability, and equity in these institutions is thus what is necessary to realize SDG 4 and SDG 5. The research needs to be developed in longitudinal and policy-evaluation designs to further develop these findings.

Works Cited

- Rahman, P. (2024). A reflexive thematic analysis exploring the experiences of undergraduate women in STEM in Bangladesh. *Discover Education*, 3(1).
<https://doi.org/10.1007/s44217-024-00185-9>
- Sarder, S., & Rani, E. (2025, October 1). Familial and social level inequalities in educational attainment: A qualitative study on Khulna University students.
<https://dergipark.org.tr/en/pub/usbed/issue/91777/1742004>
- Tewari, A. (2018). An exploration of experiences and academic challenges of African female student at University of Zululand, South Africa (UNIZULU). *Journal of Social and Development Sciences*, 8(3), 43–54. <https://doi.org/10.22610/jevr.v8i3.1992>
- Wahid, S., Ali, W., & Baloch, N. U. (2025). Exploring Sociocultural constraints on Females' Higher Education in Balochistan: a Mixed-Methods study. *KOMUNITAS INTERNATIONAL JOURNAL OF INDONESIAN SOCIETY AND CULTURE*, 17(2), 126–137. <https://doi.org/10.15294/komunitas.v17i2.31610>
- Amin, M., Islam, M., & Amin, H. (2020). What supports Females in Higher Education Progression? A Pakistani Public University Context. *Pakistan Journal of Gender Studies*, 20(1), 89–106. <https://doi.org/10.46568/pjgs.v20i1.39>
- Deblasio, L. (2024). Non-Inclusivity, Discrimination and the Othering of Female Students in Higher Education with Lived Experience of Trauma and Adversity. *IntechOpen*. doi: 10.5772/intechopen.112616
- Ellegaard, M., Jacobsen, D., & Johannsen, B. F. (2024). Peer feedback as a professional competence: Students' experiences of practicing feedback literacy in cross-professional settings..
- Hasan, N. M. N., & Akter, N. (2023). Determinants of academic major choice: a comparative study of business programs in Bangladesh. *International Journal of Research in Business and Social Science* (2147-4478), 12(9), 416–427.
<https://doi.org/10.20525/ijrbs.v12i9.3032>
- Hossain, M. T., Akter, S., Nishu, N. A., Khan, L., Shuha, T. T., Jahan, N., Rahman, M. M., & Khatun, M. T. (2023). The gender divide in digital competence: a cross-sectional study on university students in southwestern Bangladesh. *Frontiers in Education*, 8.
<https://doi.org/10.3389/educ.2023.1258447>
- Bhatti, F., & Taheri, E. (Eds). (2024). *Gender Inequality - Issues, Challenges and New Perspectives*. IntechOpen. doi: <https://doi.org/10.5772/intechopen.110910>
- SiDdiKy, M. R. (2020). Does soft skill development vary among the students? A gender perspective. *Turkish Journal of Education*, 9(3), 205–221.
<https://doi.org/10.19128/turje.658198>
- Sivakumar, Subitha and Venkataraman, Sivakumar and Balasubramanian, Dr.Lavanya, *Balancing Academic and Personal Life: Female Students' Experiences and Challenges in Higher Education* (November 04, 2024). Available at SSRN: <https://ssrn.com/abstract=5008871> or <http://dx.doi.org/10.2139/ssrn.5008871>

Appendix

Selected Stata Output Tables

Table A1: Descriptive Statistics—Key Variables

Variable	N	Mean	Std. Dev.	Min	Max
Age (coded)	400	2.38	0.77	1	4
Income level	400	2.76	0.88	1	4
Parental education	400	3.15	1.02	1	5
Enrollment difficulty	400	1.56	0.50	1	2
Harassment frequency	400	2.18	0.93	1	4
Socio-cultural restrictions	400	0.41	0.49	0	1
Satisfaction level	400	3.36	0.84	1	5

Table A2: Frequency Distribution—Age Groups

Age Category	Frequency	Percentage	Cumulative %
Below 20	58	14.5	14.5
20-22	146	36.5	51.0
23-25	183	45.5	96.5
Above 25	14	3.5	100.0
Total	400	100.0	

Table A3: Frequency Distribution—Household Income

Income Range (BDT)	Frequency	Percentage	Cumulative %
Below 15,000	27	6.75	6.75
15,001-30,000	104	26.0	32.75
30,001-50,000	192	48.0	80.75
Above 50,000	77	19.25	100.0
Total	400	100.0	

Table A4: Frequency Distribution—Harassment Experience

Harassment Frequency	Frequency	Percentage	Cumulative %
Never	50	12.5	12.5
Rarely	178	44.5	57.0
Occasionally	122	30.5	87.5
Frequently	50	12.5	100.0
Total	400	100.0	

Table A5: Cross-Tabulation—Income × Enrollment Difficulty

	Enrollment Difficulty			
Income Level	Yes	No	Total	% Difficulty
Below 15,000	21	6	27	77.8%
15,001-30,000	68	36	104	65.4%
30,001-50,000	103	89	192	53.6%
Above 50,000	33	44	77	42.9%
Total	225	175	400	56.25%

Chi² = 22.75, p < 0.001 (highly significant)

Interdisciplinary Potential of Cellular and Acellular Therapies in Cardiac Regenerative Medicine

By James Cooper

Abstract

Cardiac disease remains one of the leading causes of death worldwide, and often requires highly invasive surgical procedures. However, regenerative medicine—a relatively novel scientific field focused on developing cellular and acellular therapies—is primed to develop new surgical techniques that are far more effective and less invasive. Currently, two major subsections are being explored by researchers, including cell-based and cell-free therapies. Cellular strategies primarily focus on engrafting cardiomyocytes (CMs) directly into damaged cardiac tissue to promote regeneration and suppress left ventricular remodeling (Tenreiro 1). Additionally, acellular therapies focus on using biomaterials—either of synthetic or biologic origin—that promote cell proliferation, increase tissue engagement, and provide cell architecture. This literature review, however, encourages researchers to utilize an interdisciplinary approach that addresses the current limitations. As such, acellular biomaterials can be used as a specialized delivery mechanism for cell-based therapies. Additionally, studies have shown that cellular therapies integrated using biomaterials will engraft more effectively with native cardiac tissue, thus increasing proliferation and translatability. Accordingly, both cellular and acellular therapies curb the consequences of myocardial infarction—but in tandem, they present significantly more effective and applicable treatment options, which hold the potential to increase the quality of life for millions of patients suffering from cardiovascular disease.

Keywords cardiac regenerative medicine, acellular therapies, cellular therapies, myocardial infarction, cardiomyocytes, induced pluripotent stem cells, biomaterials.

1. Introduction

Cardiovascular disease has consistently been the leading cause of mortality worldwide, and yet, no effective non-invasive treatments have been developed to date. Every year, around 83.6 million people in the United States suffer from some form of cardiac disease—the most common presentations being myocardial infarction (MI) and peripheral artery disease (PAD) (Ungerleider 1). Both of these conditions often require significant surgical operations, such as heart transplantation or left ventricular assist devices (LVADs). These procedures—while effective—are limited by their invasive approach, which is not applicable for many patients. In addition, there is currently a serious shortage of heart donations in the United States, making this an ever more serious dilemma that calls for a less invasive, more attainable, and equally effective treatment option (Goto 2). Myocardial infarction occurs after substantial cell apoptosis, which can further lead to heart failure without medical intervention. Peripheral artery disease is primarily caused by a series of blockages in the extremities. Both of these conditions have long-term contributing factors—such as atherosclerosis—which occurs when a significant amount of plaque builds up in the arteries, causing blockages that strain the circulatory system

and prevent normal blood flow. However, new approaches in the field of cardiac regenerative medicine are being actively explored to potentially mitigate left ventricular remodeling, provide stability to damaged tissue, and allow for cardiomyocyte regeneration (OHSU 1). A variety of different therapies are in development, mainly qualifying as either cell-based or cell-free. Cellular therapies focus primarily on reinvigorating cell density in areas with damaged heart tissue. Engineering techniques aim to create cardiac patches or decellularized scaffolds from cardiomyocyte-derived induced pluripotent stem cells. Acellular therapies use a variety of chemical compounds that have shown to provide increased stability, and potentially allow for more effective cell engraftment (Bartunek 1).

This review aims to explore both cellular and acellular techniques, including their current progress and challenges, while highlighting interdisciplinary potential. The most effective solutions for mitigating the consequences of cardiovascular disease may lie in a combined approach that seeks to integrate both cell-based and cell-free therapies to improve the patient's overall chances of recovery. This review synthesized data from several peer-reviewed research articles, literature reviews, and scientific publication form within the last five years.

2. Acellular Regenerative Therapies

Acellular therapies are those that focus on using biologically compatible materials to aid in delivery, provide stability, and curb potential post-myocardial infarction complications (i.e., left ventricular remodeling).

Hydrogels are biologically or synthetically produced materials that provide provisional frameworks that aim to facilitate engraftment, angiogenesis, and increased cell density (Ungerleider 2). Studies have been conducted that explore the benefits of using hydrogels for treating myocardial infarction; however, many of these studies were conducted in small animal models, which cannot fully simulate the complex physiology of the human body and therefore, only supply proof of concept. The studies conducted in small animals have found, however, that using injectable hydrogels increases retention. In addition, other studies have posited that injecting hydrogels has led to increased systolic and diastolic function and promoted the generation of new capillaries (2). Surgeons can inject the hydrogels directly into the superficial heart tissue; therefore, this approach is relatively non-invasive compared to heart transplantation. Additionally, hydrogels can also be equipped with growth factors that further promote cell regeneration (Vasanthan 5). Furthermore, they easily integrate themselves into tissue due to their semi-liquid chemical nature. A notable model conducted on mice found that injectable hydrogels led to improved ejection function, reduced infarction damage, and promoted novel vasculature formation (5). However, there have been some concerns regarding potential immunogenicity. Some studies have shown that hydrogels that were allogeneically derived from animal sources posed potential adverse immune responses, due to the introduction of foreign animal proteins. The potential consequences can be prevented by using human-derived extracellular matrices (5).

Decellularized extracellular matrices are biologically originated structural frameworks composed of various proteins and cell structures. In regenerative medicine, this property has

been adapted by using ECMs to provide support after myocardial infarction. The cells that occupied the matrix are cleaned away with a series of chemical processes, thus allowing researchers to seed tissues with new cell types (Vasanthan 2). Additionally, the risk of an immune response is significantly reduced. Extracellular matrices are primarily composed of collagen, a generally homogenous protein structure (2). Moreover, they contain naturally occurring growth factors that exploit regenerative mechanisms, thereby providing a framework for enhanced tissue regeneration. The clinical use of decellularized extracellular matrices has shown promising results, including neoangiogenesis, reduced left ventricular remodeling, and increased engraftment with native cell tissue (3).

3. Cellular Regenerative Therapies

Cell-based or cellular regenerative therapies refer to treatments that aim to directly integrate new cells into damaged heart tissue.

Pluripotent stem cells (PSCs) are lines of cells that possess a wide range of differentiation capabilities. PSCs can specialize into nearly any type of cell, except gametes or cells of hematopoietic origin. PSCs are actively being researched, particularly in regenerative medicine for their potential clinical applications. There are two current methods used to harvest these cells, such as extracting them directly from human embryos. However, many ethical concerns have been espoused regarding this extraction method, and therefore, researchers primarily use a novel method that forgoes potential ethical boundaries. Induced pluripotent stem cells (iPSCs) are produced by restoring somatic cells to their undifferentiated form via genetic manipulation (Tenreiro 2). However, significant challenges still need to be overcome in order to effectively use iPSC-CMs in patients. In particular, iPSCs are profoundly underdeveloped when compared to adult cardiomyocytes. This is a consequence of their maturation in vitro, which fails to adequately simulate the environment needed to produce cell lines that resemble mature cardiomyocytes. Researchers have found issues such as asynchronous beat patterns and onset electrical impulses that can lead to arrhythmias. Currently, scientists have not yet produced a line of iPSC-CMs that exhibit the same level of maturity as adult cardiomyocytes (2). In addition, producing cell lines on 2D plates exacerbates the complexity issue and hinders translatability. Other approaches, such as constructing 3D structures, such as spheroids or organoids have been used to increase the maturity and complexity of cell cultures. These 3D fabrication techniques have yielded promising results, showing increased cellular communication and improved electrical stimulation (4). However, there are other ethical quandaries that call for continued research and mitigation strategies, such as preventing immunogenicity and tumorigenesis. Additionally, other concerns have been posited regarding potential genetic abnormalities, and the development of pathogenic tumor cells (Goto 3). Despite the ongoing clinical translation setbacks, researchers have continued to make significant progress. A study conducted in 2013 seeded hiPSCs into a decellularized mouse ECM. Even more promising, hiPSCs were successfully integrated into a decellularized human heart matrix. However, it should be noted

that while rudimentary recellularization may have been achieved, much more research is needed before a truly functioning human heart is developed (3).

Another intriguing application of regenerative medicine has been used in bioprinting—a process that involves using a 3D bioprinter to construct complex vasculature with cell lines that would otherwise be impossible to replicate by hand. The cells and biomaterials are known as “bioinks,” and multiple mechanisms of cell integration are currently being explored to maximize effectiveness and proliferation. One such method, known as “Freeform Reversible Embedding of Suspended Hydrogels,” extrudes bioinks from modifiable gels, which allows for the assembly of more intricate 3D structures (Tenreiro 7). This method is a promising step towards developing more translatable cell cultures, and augmenting the potential impact of regenerative technology. However, the cells produced using this method only maintained unified functionality for one day in vitro. This is also true for many other currently used bioprinting techniques, which exhibit initial functionality but eventually become asynchronous (7). Overall, while this technology shows promise, before it can be clinically translated many improvements are required to ensure effectiveness, and address safety concerns.

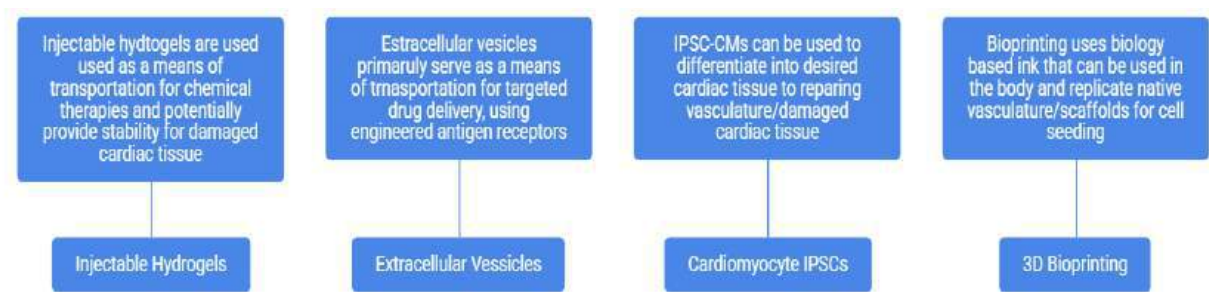


Fig 1: illustrates a brief breakdown of commonly used acellular and cellular techniques, providing a brief preface on potential pre-clinical capabilities

Cardiac Regeneration Technique	Major Results Expected	Current Progress
Injectable Hydrogels	Increased cardiac stability, reduced left ventricular remodeling after myocardial infarction, and targeted drug delivery	- Injectable hydrogels are continuing to show promising results in small animals - Few studies have translated this technology into larger animals

Extracellular Vesicles	Genetically targeted drug delivery to cardiac tissue, thus reducing potential side effects	-There are currently no FDA-approved biomaterials for the treatment of myocardial infarction. Mainly being tested in animal models, a few are undergoing clinical trials
Cardiomyocyte iPSCs	Increased cardiac stability, increased cardiomyocyte regeneration in the cardiac niche, and decreased left ventricular remodeling	- Early pre-clinical phases, shown to increase cardiac function in porcine hearts (large animal models)
3D Bioprinting	Scaffold/extracellular matrix development to increase stability and offer a seeding platform for cell-based therapies	-Mostly still in the early developmental stages, limited animal studies -Many issues regarding neovascularization and simulating the complexity of human physiology

Fig 2: shows common acellular and cellular therapies and their current clinical status

4. Interdisciplinary Potential

While many improvements are needed to address the wide range of challenges for both acellular and cellular regenerative therapies, clinical applicability and effectiveness could benefit from an interdisciplinary approach. Much of the current research hones in on either an acellular or cellular technology, separate from each other. However, the solution to improving engraftment, proliferation, biochemical mechanisms, and 3D complexity may lie in a multidisciplinary approach that incorporates numerous techniques—just as the human body itself uses multiple cell types, vessels, hormones, and enzymes to carry out one simple function. Replicating this level of physiology may call for a collection of techniques. Therefore, this literature review aims to encourage increased research, specifically focusing on using acellular and cellular therapies in tandem to maximize the potentiality of these techniques.

5. Conclusion

Both cellular and acellular therapies present potential for future clinical applications by decreasing the consequences of myocardial infarction and encouraging tissue regeneration. However, the most effective and clinically applicable use of cellular and acellular therapies is via an interdisciplinary approach, which focuses on developing an all-encompassing means of integration. The clinical applications for this technology are limitless and hold the potential to profoundly reshape cardiovascular medicine, thus rendering current treatment options obsolete.

Works Cited

- Bartunek, Jozef et al. "Integrate and boost: bioscaffolds nurture the cardiac regenerative paradigm." *Stem cell research & therapy* vol. 6,1 183. 23 Sep. 2015, doi:10.1186/s13287-015-0184-0
- Goto, Takasumi, et al. "Regenerative Medicine in Cardiovascular Disease." *Science Direct*, 5 Oct. 2024, www.sciencedirect.com/science/article/pii/S2352320424001640. Accessed 10 Jan. 2026.
- Organ Health and Science University. "Heart Regeneration." *Organ Health and Science University*, www.ohsu.edu/regenerative-medicine-center/heart-regeneration. Accessed 10 Jan. 2026.
- Tenreiro, M.F., Louro, A.F., Alves, P.M. et al. Next generation of heart regenerative therapies: progress and promise of cardiac tissue engineering. *npj Regen Med* 6, 30 (2021). <https://doi.org/10.1038/s41536-021-00140-4>
- Ungerleider, Jessica L, and Karen L Christman. "Concise review: injectable biomaterials for the treatment of myocardial infarction and peripheral artery disease: translational challenges and progress." *Stem cells translational medicine* vol. 3,9 (2014): 1090-9. doi:10.5966/sctm.2014-0049
- Vasanthan, Vishnu et al. "Promoting Cardiac Regeneration and Repair Using Acellular Biomaterials." *Frontiers in bioengineering and biotechnology* vol. 8 291. 17 Apr. 2020, doi:10.3389/fbioe.2020.00291

The Evolution of Neuromorphic Computing: An Analysis of Neuromorphic Systems, Memristive Devices, and Future Challenges By Vivian Chen

Abstract

As modern data demands push the limits of traditional von Neumann architecture, the "von Neumann bottleneck" and rising energy consumption have created a need for more efficient alternatives. Neuromorphic computing offers a solution by incorporating the biological structure and function of the human brain, prioritizing parallelism and the colocation of memory and processing. This paper provides a literature review of the evolution of neuromorphic hardware. The review further explores emerging technologies, specifically memristors, and their transformative potential in "in-sensor" computation for intelligent medicine. Despite significant advantages in energy efficiency and real-time processing, the field faces substantial hurdles, including high fabrication costs, the absence of standardized software frameworks, and the mathematical "weight transport problem." This paper concludes that the future of neuromorphic computing lies not in a single, standardized architecture, but in diverse hardware and algorithms tailored to specific applications.

Introduction

The most prominent computing architecture of the modern era, the von Neumann architecture, has played a critical role in digital development. Its core principle is clear separation between a central processing unit (CPU) and memory, and it has been remarkably successful. However, current applications of AI, autonomous systems, and big data are pushing the boundaries of von Neumann architecture, revealing its limitations. The von Neumann model constantly transfers data between memory and processing, resulting in inefficiency, consuming vast amounts of energy, and generating significant heat. With further development of next-generation applications, this becomes highly problematic, as the von Neumann model cannot keep pace with the processing demands of powerful data centers.

In search of an alternative method, researchers have turned to the most efficient computational system known, the human brain. Operating on 20 watts (just enough to power a dim lightbulb), the brain performs complex perceptual, cognitive, and motor tasks that surpass even the most powerful computing systems (Schuman 3). The field of neuromorphic computing, first established by Carver Mead in the late 1980s, aims to build hardware that mimics the brain's function and structure. Since its founding, neuromorphic computing has grown to encompass a diverse array of hardware approaches (Schuman 1).

This literature review combines current research on neuromorphic computing hardware. It examines the differences between traditional and neuromorphic computing, highlights the promising aspects of emerging hardware that takes inspiration from this method, and elucidates the challenges that remain for neuromorphic architectures .

Neuromorphic computing is rapidly evolving, and its success depends on the mutual development of specialized hardware, algorithms, and software. Because different applications

demand unique performance trade-offs, a single standardized architecture is unlikely to emerge. As Kudithipudi et al. note, the field is moving toward "a range of neuromorphic hardware solutions with different characteristics based on application needs" (805), a diversity that inherently requires a corresponding variety of specialized algorithms and software frameworks to succeed.

Differences Between Traditional and Neuromorphic Computing

Neuromorphic computing is fundamentally different from traditional computing because it takes inspiration from the brain's network of neurons and synapses (IBM, "Neuromorphic Computing"). First, this literature review presents the core biological concepts that neuromorphic computing seeks to imitate.

The first key principle is the event-driven, or "spiking," nature of computation. In the brain, information is communicated and processed through electrical pulses called "spikes"—brief bursts of electrical activity in neurons. "When a neuron becomes active or 'spikes,' it triggers the release of chemical and electrical signals that travel via a network of connection points called synapses, allowing neurons to communicate with each other" (Caballar and Stryker). Implementing this approach into hardware would create a sharp difference from traditional operational, time-driven CPUs. In traditional CPUs, power is wasted due to the system always running to receive timely signals. In a neuromorphic system, silicon "neurons" remain idle until they receive a spike, or input signal. Afterwards, they will perform a relatively simple computation to determine whether to send a spike of their own. This event-driven system is how neuromorphic computation is a source of potential for energy efficiency: Power is only consumed for computation and communication that is meaningful.

These neurological mechanisms are modeled in hardware through Spiking Neural Networks (SNNs). SNNs are often considered the third generation of neural networks, after perceptrons and modern deep learning networks. Unlike its predecessors, which use continuous values to represent information, SNNs use discrete spikes to communicate, making them much more similar to biological neural networks.

Two other key principles are parallelism and colocation of memory and processing. In the human brain, there are around 86 billion neurons connected by trillions of synapses, each functioning in parallel. Furthermore, computation is not centralized but distributed. Another crucial part is that the synapses don't just connect neurons; they are sites for memory and learning. Neuromorphic engineering seeks to physically replicate the brain's architecture to overcome the limitations of traditional computing. As Schuman et al. note, the goal is to achieve "low-power, massively parallel" capabilities, creating systems that can perform in real-time and "learn in an online fashion" (2). While modern von Neumann models mimic the brain by using electrical "spikes" to transmit data, their fundamental processing method differs. In conventional systems, the Central Processing Unit (CPU) is physically isolated from memory storage, which forces data to be shuttled back and forth. In contrast, biological synapses integrate these two functions, eliminating the need for a separate storage unit. This reduces energy costs and

bypasses the "von Neumann bottleneck," where the speed of data transfer between units limits overall performance (Caballar and Stryker). By merging memory and computation into a single structure, neuromorphic designs aim to solve the twin crises of energy inefficiency and processing latency found in standard hardware.

Different Neuromorphic Systems

The pursuit of brain-inspired computation has led to a variety of hardware designs, each with trade-offs regarding energy efficiency, computational speed, and flexibility. Steve Furber categorizes and compares several large-scale neuromorphic systems, illustrating that the field is characterized by architectural diversity rather than a single, unified approach.

Digital neuromorphic systems use binary, synchronous logic similar to traditional computers, but in a massively parallel, event-driven manner. It still uses time-based signals and 0s and 1s (on and off switches). What makes them different is that computation is spread across thousands of cores, and these cores are turned off until they receive a spike. Having similar logic allows engineers to use familiar, preexisting material.

A leading example of this approach is Spiking Neural Network Architecture (SpiNNaker), a project led by Furber himself. SpiNNaker is a massively parallel computing platform composed of over half a million ARM processor cores, each capable of simulating a population of spiking neurons (Furber). The benefit of the SpiNNaker is that it is very flexible. The neural networks are not integrated into hardware but rather modeled through software. This is because the ARM processor cores need to be programmed in order to function. This allows researchers to easily experiment with different models and configurations. The system is designed to operate similarly to biological time, making it a powerful tool for large-scale simulations of brain regions (Furber 12).

In contrast, IBM's TrueNorth chip represents a more specialized digital architecture. Described as a "neurosynaptic chip," TrueNorth was designed for extreme power efficiency in real-world cognitive applications ("TrueNorth"). Its architecture contains a network of 1 million programmable neurons and 256 million configurable synapses on a single chip. TrueNorth cannot be reprogrammed, which makes it more rigid than SpiNNaker, but it is incredibly efficient. Furber notes that TrueNorth achieves remarkable energy efficiency of 25 picojoules per synaptic operation thanks to its advanced 28-nanometer digital process and highly optimized, event-driven design (Furber 12).

Intel's Loihi chip is another prominent digital platform that emphasizes on-chip learning. As Davies et al. explain, "Loihi's architecture is optimized for sparse and nonbatched computation... placing Loihi in a diametrically opposite architectural regime compared to state-of-the-art von Neumann processors and deep learning accelerators" (915). This design allows Loihi to process data efficiently, where only a small percentage of neurons are active at any given time, a common characteristic of both biological brains and SNNs. The Loihi chip is much more flexible because it allows for "on-chip" learning which means that adaptation can be made on the hardware chip in real time (Davies et al. 913). However, its "off-chip mesh links"

have “roughly 30× lower bandwidth than its on-chip links” (Davies et al. 917). This means that neuron cores on the same chip perform at high speeds while communications between multiple chips require much more time.

Analog neuromorphic systems use the physical properties of electronic components, such as the current flowing through a transistor or the charge on a capacitor, to directly emulate the continuous dynamics of biological neurons and synapses. These systems can offer unparalleled speed and energy efficiency, but often at the cost of programmability and susceptibility to electronic noise and device variation. The Stanford Neurogrid uses sub-threshold analog circuits to emulate the ion channels and membrane potentials of biological neurons with high fidelity (Furber 4). This makes it an excellent tool for computational neuroscience, allowing researchers to build and test detailed models of neural circuits. It operates in biological real-time and consumes very little power, making it suitable for potential applications in low-power interfaces.

The BrainScaleS system from Heidelberg University takes a different approach, using wafer-scale integration of analog electronics to operate at remarkable speeds—10,000 times faster than biological real-time (Furber 4). This “accelerated” operation is not for real-time interaction with the world but is a powerful feature for scientific research. It allows neuroscientists and computer scientists to run long-term learning experiments or sweep through large parameter spaces for network optimization in a matter of minutes or hours, rather than days or weeks.

While most current large-scale systems are built with conventional silicon CMOS technology, a vibrant area of research explores novel materials and devices that can more directly and efficiently emulate neural components. Caballar and Stryker point out that “researchers are also looking into new types of materials, such as ferroelectric and phase-change materials,” to overcome the limitations of standard CMOS built with the von Neumann model.

A primary application domain is at the “edge,” where data is generated by sensors and must be processed immediately without the latency and power cost of sending it to the cloud. Neuromorphic vision sensors, for example, only output data when a pixel detects a change in brightness, generating sparse, efficient data streams. Coupled with a neuromorphic processor like TrueNorth or Loihi, these systems can perform always-on object recognition, gesture tracking, or anomaly detection for surveillance with milliwatt-level power consumption, enabling extended operation time while taking up less space. Systems like SpiNNaker and BrainScaleS were built with a primary goal of simulating brain function. They allow neuroscientists to create and test large-scale models of brain regions, such as the cortex or basal ganglia, to better understand neural coding, learning, and disorders. Beyond neuroscience, the massive parallelism of these systems can be applied to other complex scientific problems, such as solving optimization problems or simulating other natural phenomena.

Memristors

A prominent emerging device is the memristor (a portmanteau of “memory” and “resistor”). A memristor is a non-volatile electronic memory element whose resistance changes

based on the history of the electrical charge that has flowed through it. This "memory" of past activity is strikingly similar to the behavior of a biological synapse, whose strength changes based on the timing of pre- and post-synaptic spikes. This makes memristors an ideal candidate for physically realizing the colocation of memory and processing (IBM, "Neuromorphic Computing").

Research into memristors is advancing rapidly. Kumar et al. describe a hierarchy of memristive complexity, from first-order devices that provide simple synaptic plasticity to third-order devices that exhibit neuron-like dynamics, enabling higher-complexity neuromorphic computing. These devices hold the promise of creating ultra-dense, low-power synaptic arrays that could dramatically increase the scale and reduce the energy consumption of future neuromorphic systems. Furber also acknowledges this potential, noting that although these novel devices "do not feature in current large-scale projects, they may well play a major role in future neuromorphic systems" (14). One way that memristors have started to play a role in our lives is in the medical field.

In "Memristor-based Neuromorphic Diagnostic Systems: an emerging strategy for intelligent medicine," Sun et al. research the possible applications of memristors in medicine. Recent breakthroughs have transferred memristors from being purely theoretical to becoming promising key components in biosensing, intelligent electronic skin (e-skin), neuromorphic diagnosis and treatment systems. Sun et al. have proposed that the resistance dynamic qualities of memristors, "can be well used in information memory, also called resistance random access memory (RRAM)." Memristors have a structure very similar to a synapse since they both consist of a "sandwich" like structure. Memristors are able to achieve a "sandwich" like structure through a layer of semiconductor and insulator sandwiched between two conductive electrodes. This similarity leads to "natural structural advantage in the preparation of implantable neuromorphic diagnostic and therapeutic systems" (Sun et al. 9), largely because the simple design allows for miniaturization. This structural simplicity further enables the organization of memristors into high-density crossbar arrays, where thousands of "sandwiches" are mapped into a grid. This configuration allows for massive parallelism in sensing, mimicking the way biological systems process sensory input. Unlike traditional sensors that must shuttle raw data to a central processor, these neuromorphic arrays can perform "in-sensor" computation, filtering and decoding complex biological signals in the same place. This localized processing not only increases sensing resolution but also minimizes the latency and power consumption typically associated with high-bandwidth medical monitoring (Sun et al. 2). Additionally, unlike traditional sensors, which require a constant flow of electricity to "hold" a thought, a memristor retains its state even when power is cut making it non-volatile. In a medical implant (like a pacemaker or a neural decoder), this means the device only consumes power when it is actively processing a signal, drastically extending the interval between battery replacements.

Challenges

The unique properties of neuromorphic hardware, low power, real-time processing, and event-driven computation, make it ideally suited for a class of applications where traditional computing struggles. Despite its promising nature, the large-scale adoption of neuromorphic computing faces many challenges that need to be overcome.

A critical barrier is the high cost of manufacturing complex neuromorphic chips. Davies et al. provide a clear economic explanation for this, noting that "The per-bit cost of monolithic dynamic random access memory (DRAM) is on the order of $100\times$ cheaper than the per-bit cost of the densest memory available in a logic process" (922). Traditional CPUs can partition their design between cheap DRAM and expensive logic circuits. In contrast, a neuromorphic architecture like Loihi, which shares memory and processing, must be fabricated in an expensive process. This makes scaling up the technology costly for mainstream applications. As Davies et al. have stated, "Without breakthroughs in the density of low-energy, high-speed CMOS-integrated memory, the cost will limit the proliferation of all but small-scale neuromorphic technology" (923). This makes scaling up the technology prohibitively expensive for mainstream applications.

However, hardware is only one part of the equation. Programming these non-Von Neumann systems is difficult, and there is a steep learning curve for developers accustomed to traditional programming. Kudithipudi et al. identify that creating strong software frameworks is crucial "to minimize unnecessary duplication of existing components and to facilitate transitions between ecosystems" (803). The field currently lacks the standardized programming interfaces and mature toolchains that enabled the success of traditional systems. Furthermore, providing a benchmark for neuromorphic systems against traditional AI is challenging. This is due to its strength being energy efficiency and event-based events rather than sheer speed of getting tasks done.

Beyond the difficulty of programming and the high cost of logic-based memory, the algorithms used to train these systems present a significant efficiency hurdle. Modern AI relies heavily on backpropagation. Backpropagation is a mathematical method used to calculate the "gradient" or error of a network to determine how synaptic weights should be adjusted (Bergmann and Stryker). However, backpropagation is at odds with the event-driven, local nature of spiking neural networks. As Bergmann and Stryker explain, backpropagation requires a "backward pass" that relies on global information from the entire network to function. This creates what Dampfhooffer et al. define as a "weight transport problem," where data must be constantly shuttled across the chip to update synapses (Dampfhooffer et al. 11910). For a system designed to eliminate the von Neumann bottleneck through the colocation of memory and processing, this massive data movement reintroduces a version of the bottleneck at the algorithmic level. Furthermore, because SNNs communicate through discrete, sudden "spikes" rather than continuous values, traditional backpropagation—which relies on smooth calculus and derivatives—is mathematically difficult to implement without complex, energy-intensive workarounds (Dampfhooffer et al. 11912).

For emerging technologies like memristors, significant hurdles remain in device reliability, endurance, and variability. Creating large-scale, functional crossbar arrays of memristors without significant yield issues is a major focus of materials science and electrical engineering research. Integrating these novel materials seamlessly with standard CMOS logic also presents a fabrication challenge.

Neuromorphic computing represents a fundamental rethinking of computer architecture, driven by the need for greater efficiency and new computational capabilities. Ultimately, experts suggest that there will not be a "one-size-fits-all solution" for NC systems. Instead, the focus should shift to investigating diverse capabilities to understand which features are essential for different application needs (Kudithipudi et al.). The challenge now lies in developing robust frameworks and hardware to minimize duplication and facilitate the transition between distinct architectural ecosystems (Kudithipudi et al.).

Works Cited

- Bergmann, Dave, and Cole Stryker. "What is Backpropagation?" IBM Think, IBM, 2024, www.ibm.com/think/topics/backpropagation. Accessed 11 December. 2025
- Caballar, Rina Diane, and Cole Stryker. "Neuromorphic computing." *What is neuromorphic computing?*, 2025, www.ibm.com/think/topics/neuromorphic-computing. Accessed 15 Oct. 2025.
- M. Dampfhofer, T. Mesquida, A. Valentian, and L. Anghel, "Backpropagation-Based Learning Techniques for Deep Spiking Neural Networks: A Survey," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 9, pp. 11906-11921, Sept. 2024, doi: 10.1109/TNNLS.2023.3263008.
- Davies, M. et al. "Advancing Neuromorphic Computing With Loihi: A Survey of Results and Outlook." *Proceedings of the IEEE*, vol. 109, no. 5, May 2021, pp. 911-934. *IEEE Xplore*, doi:10.1109/JPROC.2021.3067593.
- Furber, Steve. "Large-Scale Neuromorphic Computing Systems." *Journal of Neural Engineering*, vol. 13, no. 5, 2016, p. 051001. *IOPscience*, doi:10.1088/1741-2560/13/5/051001.
- Kudithipudi, D., et al. "Neuromorphic Computing at Scale." *Nature*, vol. 637, 2025, pp. 801–812. *Nature*, doi:10.1038/s41586-024-08253-8.
- Open Neuromorphic. "TrueNorth." 2023, open-neuromorphic.org/neuromorphic-computing/hardware/truenorth-ibm/. Accessed 15 Oct. 2025.
- Schuman, Catherine D., et al. "A Survey of Neuromorphic Computing and Neural Networks in Hardware." *arXiv*, 2017, arXiv:1705.06963.
- Sukel, Kayt. *Neurotransmission: The Synapse*. dana.org/app/uploads/2023/09/fact-sheet-neurotransmission-synapse-baw-2020.-1.pdf.
- Sun B, Zhang J, Yang C, et al. Memristor-based Neuromorphic Diagnostic Systems: an emerging strategy for intelligent medicine. *ACS Nano Medicine*. Published online November 10, 2025. doi:10.1021/acsnanomed.5c00084

Fourier-Domain Feature Extraction for Stable Sensory Feedback in Neuro-Haptic Prosthetics By Martin Meneses

Abstract

Restoring meaningful sensory feedback remains a fundamental challenge in modern prosthetic systems. While electromyographic (EMG) signals are widely used for prosthetic control, time-domain interpretations of these signals are highly sensitive to noise, motion artifacts, and transient fluctuations, resulting in unstable and unintuitive haptic feedback. This paper proposes a frequency-domain approach to sensory feedback generation based on Fourier-domain feature extraction. By transforming EMG signals into the frequency domain and computing band-specific spectral power, sustained patterns of muscle activation can be isolated while suppressing noise and spurious disturbances. A representative numerical example demonstrates how band power provides a smoother and more physiologically aligned control signal than raw amplitude measures. The paper further illustrates how lightweight artificial intelligence models can operate on frequency-domain features to adapt sensory mappings across users without sacrificing interpretability or stability. Together, these results show that Fourier-domain feature extraction offers a robust mathematical foundation for neuro-haptic prosthetics and enables controlled integration of adaptive learning for personalized sensation.

1. Introduction

Modern prosthetic systems have achieved significant progress in mechanical dexterity, yet the restoration of meaningful sensation remains a central challenge (Hudgins, Parker, and Scott). While users can generate movement through electromyographic signals, the lack of stable and intuitive sensory feedback limits precision and embodiment.

Electromyography, the study of the very muscle-electric signals (EMG) needed to drive an autonomous hand, is widely used in prosthetic control because it provides a non-invasive measure of muscle activation (Merletti and Parker). However, EMG signals are inherently noisy and highly variable, making direct interpretation difficult. When analyzed in the time domain, these signals often fail to provide consistent information needed for fine-grained sensory feedback (Farina, Merletti, and Enoka).

Mathematical tools such as Fourier analysis offer a way to extract structure from EMG signals without increasing system complexity (Smith). By transforming EMG signals into the frequency domain, underlying patterns of muscle activity can be isolated and used to generate more stable sensory representations. This paper explores how frequency-domain analysis can improve sensory feedback in neuro-haptic prosthetics and outlines how artificial intelligence may further enhance these systems.

2. Background

2.1 EMG Signals in Prosthetic Control

Electromyographic signals originate from the electrical activity produced by motor units during muscle contraction (Merletti and Parker). Surface electrodes placed on the skin detect these voltage fluctuations and translate them into control signals for prosthetic devices. This approach is widely used due to its non-invasive nature and relatively simple hardware requirements.

However, EMG signals provide only an indirect representation of user intent. The recorded signal reflects the combined activity of multiple muscles, is sensitive to electrode placement, and varies with fatigue and movement. While such variability can often be tolerated for basic motor commands, it presents a significant challenge for generating consistent sensory feedback.

Unlike motor control, which typically relies on discrete activation thresholds, sensory feedback requires continuous and stable signal interpretation. Small fluctuations in EMG amplitude can lead to disproportionate changes in haptic output, resulting in jittery or unintuitive sensations. This limitation motivates the need for representations that capture sustained patterns of muscle activity rather than instantaneous signal changes.

2.2 The Sensory Feedback Bottleneck

Time-domain EMG analysis represents muscle activity as a sequence of voltage amplitudes over time. While this representation is sufficient for detecting the onset or absence of contraction, it provides limited insight into the internal structure of the signal. As a result, transient spikes, motion artifacts, and noise often appear indistinguishable from meaningful muscle activation.

For sensory feedback, this limitation is critical. Haptic systems require smooth, continuous modulation to convey pressure, texture, or force. When sensory output is driven directly by time-domain EMG amplitude, short-lived fluctuations can produce unstable or exaggerated feedback responses. Simple filtering techniques can reduce high-frequency noise, but they do not address the underlying ambiguity between sustained activation and transient disturbances.

This bottleneck arises because time-domain representations emphasize instantaneous signal magnitude rather than persistent patterns of muscle activity. To generate stable sensory feedback, prosthetic systems require representations that capture how energy is distributed within the signal over time. This motivates the use of frequency-domain analysis, which provides a structured view of EMG signals beyond raw amplitude variation.

2.3 Fourier Transform: Conceptual Background

The Fourier Transform (FT), developed by Jean-Baptiste Joseph Fourier (1768–1830), is a mathematical tool that converts a time-based function into its representation in the complex frequency domain (3Blue1Brown). The Fourier Transform is widely used in signal and circuit analysis since every signal, including audio, is composed of multiple frequencies. It provides a way to represent the signal by separating it into its individual frequency components (Gray). By

applying Euler's identity, it expresses signals as combinations of sinusoids, shifting the variable from time 't' to frequency 'ω'. This process is commonly written as $F[f(t)] = F(w)$.

$$F(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega t} f(t) dt = \int_{-\infty}^{+\infty} (\cos\omega t - i\sin\omega t) f(t) dt$$

The FT represents a signal in terms of its frequency content rather than its variation over time. Instead of describing how a signal changes moment by moment, it describes how strongly different frequency components contribute to the overall signal

In practical terms, the Fourier Transform decomposes a time-domain signal into a spectrum that shows how signal energy is distributed across frequencies. This representation is especially useful for biological signals such as EMG, where meaningful muscle activity is often embedded within specific frequency ranges while noise and artifacts dominate others

For EMG analysis, this transformation enables sustained patterns of muscle activation to be distinguished from transient fluctuations. By focusing on frequency-domain features rather than instantaneous amplitudes, prosthetic systems can extract more stable representations of user intent, which are better suited for generating continuous sensory feedback.

3. Signal Detection and Frequency-Domain Feature Extraction

Electromyographic signals are acquired as continuous voltage measurements and digitized through analog-to-digital conversion. Once sampled, the EMG signal is represented as a discrete-time sequence suitable for computational analysis. This discretization enables signal transformations that reveal structure not apparent in raw voltage traces.

To analyze EMG signals in the frequency domain, short segments of the sampled signal are processed using a discrete Fourier transform, typically implemented via the Fast Fourier Transform (FFT) (Smith). Applying the transform over short time windows preserves temporal locality while allowing frequency content to be examined. The result is a spectrum that quantifies the contribution of each frequency component within the window.

Feature extraction is performed by analyzing the distribution of spectral energy across selected frequency bands (Farina, Merletti, and Enoka). Rather than relying on instantaneous amplitude, these features summarize sustained muscle activity over time. Band-specific energy measures reduce sensitivity to transient noise and provide a more stable basis for mapping muscle activation to sensory feedback. These frequency-domain features form the input to downstream control or interpretation stages.

4. Representative Frequency-Domain Calculation

To translate raw physiological activity into meaningful sensory feedback, the system moves beyond simple amplitude tracking. Instead, it utilizes frequency-domain analysis to extract stable features that correlate with motor unit recruitment patterns. To make the process of frequency domain feature extraction concrete, a simplified numerical example is presented.

While practical prosthetic systems typically employ longer windows (128 or 256 samples), a short example reveals the mechanics of the transformation without obscuring the underlying principles. The dataset presented here is illustrative and intended to demonstrate feature separability rather than report clinical measurements. After a segment of EMG data is digitized at a sampling rate f_s , it is divided into a time window of N samples. The Discrete Fourier Transform (DFT) is then applied to decompose the time-series signal into its constituent frequency components:

$$X[k] = \sum_{n=0}^{N-1} x[n] * e^{-j\frac{2\pi}{N}nk}$$

$$k = 0, 1, 2, 3, \dots N - 1$$

Where:

$x[n]$ is the n -th sample of the EMG signal in the time domain

$X[k]$ is the k -th frequency component (the "bin")

N is the total number of samples in the window

$e^{-j\frac{2\pi}{N}nk}$ represents the complex sinusoids used to “probe” the signal specific frequencies

Consider a short window of EMG data sampled at 1000 Hz, represented by four discrete samples:

$$x[0] = 10 \quad x[1] = -2 \quad x[2] = 8 \quad x[3] = 4$$

Calculating the DC Component ($k = 0$)

The first bin represents the average signal offset (0 Hz).

$$X[0] = 10 + (-2) + 8 + 4 = 20$$

This component is typically discarded in EMG-based sensory feedback, as it reflects baseline offset rather than muscle activation dynamics.

Calculating the Fundamental Frequency ($k = 1$)

This bin represents the energy at 250 Hz (calculated as $\frac{f_s}{N}$). Using Euler’s formula:

$$X[1] = 10e^0 + (-2)e^{-j\frac{\pi}{2}} + 8e^{-j\pi} + 4e^{-j\frac{3\pi}{2}}$$

Using standard trigonometric identities:

$$e^{-j\frac{\pi}{2}} = -j \quad e^{-j\pi} = -1 \quad e^{-j\frac{3\pi}{2}} = j$$

$$X[1] = 10 + 2j - 8 + 4j = 2 + 6j$$

Determining Magnitude and Band Power

To drive a haptic actuator, we require the magnitude $|X[k]|$ and power (P). Magnitude is calculated using:

$$|X[k]| = \sqrt{Re(X[k])^2 + Im(X[k])^2}$$

For the 250 Hz component:

$$|X[1]| = \sqrt{2^2 + 6^2} = \sqrt{40} \approx 6.32$$

The corresponding spectral power is:

$$P_{250hz} = P_1 = (|X[1]|)^2 = 40$$

While individual bins provide high resolution, practical EMG analysis focuses on Band Power $P(band)$. By summing the squared magnitudes within a physiologically relevant range (50–150 Hz), the system captures the aggregate energy of muscle activation while ignoring high-frequency noise and low-frequency motion artifacts.

This aggregation acts as a natural "smoothing" mechanism. Because Band Power represents the total energy over a window of time, a single noisy spike in one sample does not radically shift the output, resulting in a more stable control signal than raw peak-to-peak amplitude.

In a neuro-haptic prosthetic system, this power value contributes to the sensory feedback command, such as the vibration intensity of a piezoelectric actuator or the pressure of a pneumatic cuff. Suppose the maximum expected band power during full muscle contraction is $P(max) = 200$, and the haptic actuator is driven by an 8-bit pulse-width modulation (PWM) signal ranging from 0 to 255. A linear mapping can be defined as:

$$Intensity_{haptic} = clamp(\frac{P(band)}{P(max)} * 255, 0, 255)$$

For the given example sample worked above, substitute the computed values:

$$Intensity_{haptic} = \frac{40}{200} * 255 = 51$$

This value determines the vibration or pressure intensity delivered to the user.

Signal State	EMG Feature $P(\text{band})$	Haptic Output	User Perception
Rest	Near Zero	No Vibration	No Sensation
Light Grip	Low Spectral Energy	Low Frequency/Amplitude	Soft Touch
Strong Contraction	High Spectral Energy	High Frequency/Amplitude	Firm Pressure

Table 1. General frequency-domains feature clusters corresponding to different sensory states.

This example illustrates why frequency-domain features improve sensory stability. Because band power is computed as a summation over multiple frequency components, isolated noise spikes in individual samples or bins have limited influence on the final control signal. In contrast to time-domain amplitude mappings, this aggregation smooths the sensory response and prevents abrupt or uncomfortable feedback artifacts.

Furthermore, the ability to selectively ignore low-frequency components, such as the DC offset, prevents spurious sensations caused by electrode bias or slow drift. As a result, frequency-domain feature extraction provides a robust and physiologically meaningful basis for neuro-haptic sensory feedback (Smith).

5. Discussion

The results of the representative calculation highlight why frequency-domain feature extraction is better suited for sensory feedback than time-domain amplitude measures. By aggregating spectral energy over a physiologically relevant frequency band, band power captures sustained muscle activation rather than instantaneous fluctuations. This aggregation inherently suppresses the influence of short-lived disturbances such as motion artifacts, electrode noise, or electrical interference.

For haptic feedback, stability is not optional. Unlike motor commands, which can tolerate threshold-based decisions, sensory perception requires smooth and continuous modulation. Time-domain features such as raw amplitude or RMS respond directly to transient spikes, producing abrupt changes in feedback intensity that users perceive as jitter or vibration instability. In contrast, frequency-domain band power evolves gradually over time, resulting in a sensory signal that more closely reflects intentional muscle effort.

Importantly, this improvement is not achieved through complex filtering or heuristic smoothing, but through a change in signal representation. By analyzing how signal energy is distributed rather than how voltage fluctuates moment to moment, frequency-domain features provide a more physiologically aligned basis for neuro-haptic sensation.

Root Mean Square (RMS) amplitude is widely used in prosthetic control due to its computational simplicity and effectiveness for basic movement detection (Hudgins, Parker, and Scott). However, RMS is a broadband metric: all components of the signal, including noise and artifacts, contribute equally to the final value. In sensory feedback applications, this characteristic directly translates into unstable output.

Frequency-domain band power, while requiring additional computation via an FFT, allows selective inclusion of meaningful signal components. By restricting analysis to an activation band (for example, 50–150 Hz), common sources of interference can be systematically excluded, including low-frequency motion artifacts and high-frequency electronic noise. The result is a smoother sensory envelope that improves transparency, defined as how naturally the user perceives the prosthetic feedback as an extension of their own body.

Despite its advantages, frequency-domain feature extraction introduces trade-offs that must be considered in real-world prosthetic systems. Window length selection affects both spectral resolution and system latency: longer windows provide more stable frequency estimates but delay feedback, while shorter windows improve responsiveness at the cost of increased variability.

Additionally, fixed frequency bands may not optimally capture muscle activation patterns for all users. Variations in physiology, electrode placement, and fatigue can shift the spectral characteristics of EMG signals. While band power improves robustness compared to time-domain features, adaptive strategies may be required to maintain optimal performance across users and conditions.

These limitations motivate the integration of higher-level adaptive methods capable of tuning parameters without compromising signal interpretability.

The frequency-domain framework established here provides a stable and physiologically meaningful representation of EMG signals. Building on this foundation, artificial intelligence techniques can be employed to adapt feature mappings across users and contexts, enhancing personalization while preserving signal structure.

5.1 Artificial Intelligence as an Extension to Frequency-Domain Features

The frequency-domain framework developed in this paper provides a stable and interpretable representation of EMG signals for neuro-haptic sensory feedback. However, while band power features substantially reduce jitter and noise sensitivity, fixed mappings between spectral energy and haptic output may not generalize optimally across users or operating conditions. Differences in muscle physiology, electrode placement, and fatigue can shift spectral characteristics over time, motivating adaptive strategies that preserve signal stability while enabling personalization.

Artificial intelligence can be integrated as a secondary interpretive layer operating exclusively on frequency-domain features. In this role, AI does not replace Fourier-based signal processing, but learns how extracted features should be mapped to user-perceived sensation. This separation ensures that the system remains robust and interpretable while allowing controlled adaptation.

5.2 Example Dataset and Feature Representation

To illustrate a practical implementation, consider a representative dataset derived from frequency-domain EMG analysis. Each data point corresponds to a short EMG window processed using the FFT and summarized through band power extraction. The feature vector includes spectral energy in multiple frequency ranges to capture both activation and noise characteristics.

Example feature set:

P_{50-150} : Primary Activation Band Power

P_{20-50} : Low-frequency band (motion artifacts)

$P_{150-300}$: High-frequency band (electronic noise)

Each sample is labeled according to the user's intended interaction state during calibration.

Sample	P_{50-150}	P_{20-50}	$P_{150-300}$	State
1	5.2	1.1	0.9	Rest
2	18.4	2.0	1.2	Light Contact
3	42.7	2.4	1.8	Firm Contact
4	6.1	1.0	1.1	Rest
5	21.9	1.8	1.0	Light Contact
6	55.3	2.6	2.1	Firm Contact

Table 2. *Dataset of frequency-domain feature clusters corresponding to different sensory states.*

As shown in Figure 2, the activation band power increases monotonically with interaction intensity, while motion- and noise-related bands remain relatively stable. This structured separation indicates that frequency-domain features provide a low-dimensional and linearly separable representation of user intent. As a result, simple classifiers can reliably distinguish interaction states without operating on raw EMG signals.

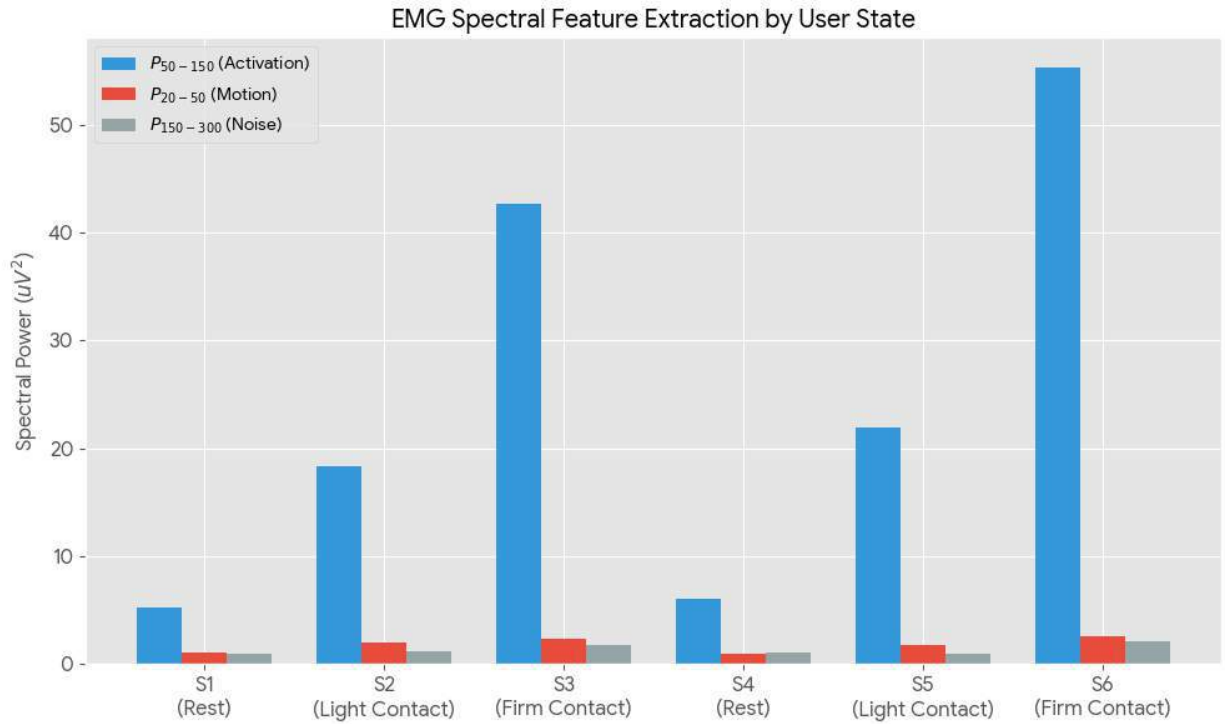


Figure 2: Spectral band power features extracted from representative EMG windows across user states. The activation band (50–150 Hz) exhibits clear separation between rest, light contact, and firm contact, while motion and noise bands remain comparatively low. This separation motivates the use of lightweight learning models for user-adaptive sensory classification.

This dataset reflects how increased muscle engagement produces higher spectral energy in the activation band while noise-related bands remain comparatively stable.

Using the extracted frequency-domain features, a lightweight supervised learning model can be trained during a short user calibration phase. In this example, a multiclass logistic regression model is used to associate spectral features with discrete sensory states such as rest, light contact, and firm contact.

The training process operates on feature vectors rather than raw EMG signals, ensuring low dimensionality and real-time feasibility. The following code snippet illustrates a representative training and inference pipeline:


```

import numpy as np
from sklearn.linear_model import LogisticRegression

# Example frequency-domain feature dataset
X = np.array([
    [5.2, 1.1, 0.9],
    [18.4, 2.0, 1.2],
    [42.7, 2.4, 1.8],
    [6.1, 1.0, 1.1],
    [21.9, 1.8, 1.0],
    [55.3, 2.6, 2.1]
])

# Corresponding sensory states
y = np.array([0, 1, 2, 0, 1, 2]) # 0: Rest, 1: Light, 2: Firm

# Train user-specific model
model = LogisticRegression(multi_class="ovr")
model.fit(X, y)

```

Figure 3: Python Code AI-Model Example Application using Sample Dataset and Feature Representation

During real-time operation, each new EMG window is transformed into the frequency domain, band power features are computed, and the trained model predicts the appropriate sensory state. This prediction is then mapped to a haptic output level, such as vibration amplitude or pressure intensity.

This feature-guided learning approach offers several advantages. First, adaptation occurs without altering the underlying signal representation, preserving stability. Second, the model remains interpretable, as predictions can be traced directly to changes in spectral energy. Third, computational requirements are modest, allowing deployment on embedded processors commonly used in prosthetic systems.

The AI component presented here is intentionally limited in scope. It serves as a practical demonstration of how learning-based methods can complement, rather than replace, mathematically grounded signal processing. This general framework, as seen in Figure 4, using more advanced tools and machinery, can be a fundamental step forward into developing AI-driven neuro-haptic prosthetics.

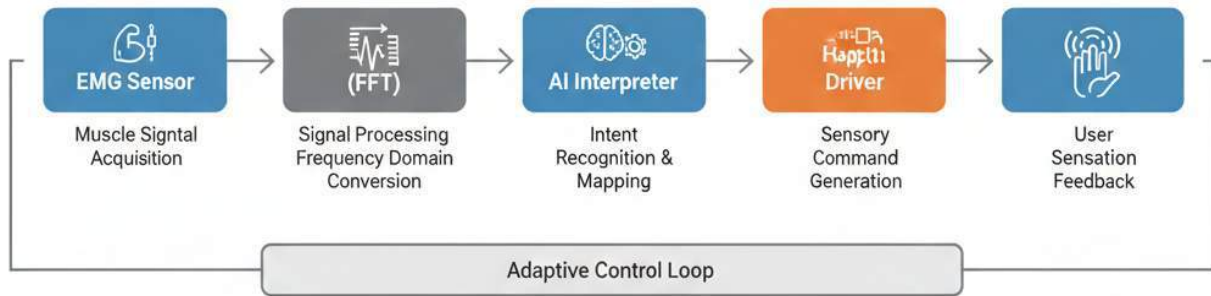


Figure 4: *Neuro-Haptic Feedback System: Simplified Data Flow*

Future work may explore continuous adaptation, user-in-the-loop calibration, and finer-grained sensory state definitions, but the core insight remains unchanged: robust sensory feedback begins with structured signal representation.

6. Conclusion

This paper set out to address a central limitation in neuro-haptic prosthetics: the instability of sensory feedback derived from time-domain EMG signals. While traditional amplitude-based features are effective for basic motor control, they are poorly suited for generating continuous and natural sensation due to their sensitivity to noise and transient disturbances. As a result, users often experience jittery or unintuitive haptic feedback that undermines precision and embodiment.

By reframing EMG interpretation in the frequency domain, this work demonstrates how Fourier-based feature extraction provides a more stable and physiologically meaningful representation of muscle activity. Band power aggregates spectral energy over time, inherently smoothing isolated fluctuations and allowing sustained motor unit recruitment patterns to drive sensory feedback. A representative calculation illustrates how this approach produces gradual, interpretable changes in haptic output, avoiding abrupt artifacts common in time-domain mappings.

Building on this stable foundation, the paper shows how artificial intelligence can be incorporated as a secondary interpretive layer operating on frequency-domain features rather than raw signals. This separation preserves robustness and transparency while enabling user-specific adaptation through lightweight, computationally efficient models. Importantly, learning is guided by structured spectral features, ensuring that adaptation enhances rather than destabilizes sensory feedback.

Overall, the results suggest that robust neuro-haptic sensation does not require increasingly complex hardware or opaque end-to-end learning systems. Instead, meaningful progress arises from mathematically grounded signal representations that align with physiological structure. Frequency-domain feature extraction provides such a foundation, enabling stable sensory feedback today and scalable, personalized prosthetic systems in future work.

Works Cited

- Balakrishnan, A. V. *Applied Functional Analysis*. Springer-Verlag, 1981.
- Cheever, Erik. “Fourier Transform Properties and Applications.” Swarthmore College Lecture Notes, Swarthmore College, 2019.
- Farina, Dario, Roberto Merletti, and Roger M. Enoka. “The Extraction of Neural Strategies from the Surface EMG.” *Journal of Applied Physiology*, vol. 96, no. 4, 2004, pp. 1486–1495. *Journal of Applied Physiology*, <https://doi.org/10.1152/japplphysiol.01070.2003>
- Gray, Robert M. “Toeplitz and Circulant Matrices: A Review.” *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, 2006, pp. 155–239.
- Hudgins, Bernard, Philip Parker, and Robert N. Scott. “A New Strategy for Multifunction Myoelectric Control.” *IEEE Transactions on Biomedical Engineering*, vol. 40, no. 1, 1993, pp. 82–94. *IEEE Xplore*, <https://doi.org/10.1109/10.204774>
- Merletti, Roberto, and Philip A. Parker. *Electromyography: Physiology, Engineering, and Non-Invasive Applications*. IEEE Press, 2004.
- Pinsky, Mark A. *Introduction to Fourier Analysis and Wavelets*. Brooks/Cole, 2002.
- Smith, Steven W. *The Scientist and Engineer’s Guide to Digital Signal Processing*. California Technical Publishing, 1997.
- 3Blue1Brown. “But What Is the Fourier Transform? A Visual Introduction.” YouTube, uploaded by 3Blue1Brown, 2018, <https://www.youtube.com/watch?v=spUNpyF58BY>

The Role of Payment for Order Flow in The Rise of Commission-Free Trading Platforms

By Julien Cohen

Abstract

This paper examines how Robinhood Markets, Inc. has leveraged payment for order flow (“PFOF”) as a core revenue-generating mechanism to sustain profitability while offering commission-free trading to retail investors. By eliminating traditional trading commissions, Robinhood disrupted the brokerage industry and catalyzed a broader shift toward zero-commission trading across the sector. However, this shift raised critical questions about the company's business sustainability and revenue model. Through a detailed analysis of Robinhood's business practices, industry standards, and regulatory complexities, this paper examines the significance of PFOF in the company's revenue structure and operational framework. The paper investigates how Robinhood monetizes trade volume by routing customer orders to market makers for execution in exchange for fees, the regulatory scrutiny and ethical implications associated with this model, and its impact on order execution quality. The paper further contextualizes Robinhood's approach within broader trends in fintech innovation and retail investor behavior. Ultimately, the paper concludes by suggesting that while PFOF has enabled Robinhood to scale rapidly and attract a new generation of investors, its ongoing dependence on PFOF puts the company under increasing regulatory watch, faces changing market conditions, and rising expectations from customers. Because of these dangers, Robinhood should shift away from the PFOF model and will need to be more open about how it operates, focus on giving users good trade execution, and work to earn investors’ long-term trust as the industry changes.

Introduction

In January 2021, a Wall Street Bets subgroup calling themselves (“Ape”) investors drove the stock prices of companies like GameStop and AMC to historic heights, challenging powerful Wall Street hedge funds. Then, on a pivotal Thursday morning, the trading app Robinhood, the very platform that promised to democratize finance, abruptly restricted its users from buying these volatile stocks, locking millions of retail traders out of the market. This sudden act, which became the subject of intense regulatory scrutiny and inspired films like *Dumb Money*, can be linked to Robinhood’s use of the Payment for Order Flow system, which is when a brokerage receives a small payment from a market maker for sending its customers stock trades to that market maker for execution, allowing the broker to offer commission free trading while the market maker profits from the bid-ask spread. The incident forced investors to ask a simple and critical question that will be explored in part later in this paper: Does the broker's revenue from Payment for Order Flow mean its loyalties are fundamentally aligned with the market makers who pay it, rather than the retail users it claims to serve?

This review paper aims to provide a comprehensive analysis of PFOF, examining its mechanics, historical context, and overall impact on the modern financial world. Part I defines

PFOF and identifies the key players involved, as well as the regulatory landscape that has allowed it to flourish. Part II explores the historical development of Robinhood Markets, Inc., as a central case study, detailing how the company's business model is built upon PFOF and the controversies the business endured along its journey. Part III analyzes how PFOF is used across the entire brokerage industry, comparing the practices of traditional brokers with those of commission-free platforms. Part IV weighs the benefits and drawbacks of PFOF. It assesses whether the practice truly serves the best interests of the retail investor by considering both the apparent advantages and the hidden costs. Ultimately, this paper argues that while PFOF has been a powerful engine for expanding access to financial markets, the use of PFOF likely does more damage to free market concepts than it does to aid them.

I. What is PFOF?

A. Definition and a Basic Explanation of Payment for Order Flow (“PFOF”)

PFOF is a common but debated financial market transaction type in the U.S. stock markets. In essence, PFOF is the money a stock brokerage firm gets for sending its customers' orders to a certain market maker to be traded. Instead of sending an order to a public stock exchange, a broker sends it to a market maker in exchange for a fee, which is usually a very small amount of money per share. Economically, PFOF monetizes the lack of information among retail traders, as these smaller, less informed orders are less risky for market makers to fill than those from institutional players. The predictability in order flow from retail traders allows market makers to easily manage their risk, complete trades for a small profit, and earn money from the difference between the buy and sell prices.

The PFOF process is like a closed circle that occurs outside of public exchanges. When a regular investor places an order to buy or sell a stock through a PFOF-enabled broker, that order is not immediately sent to a public exchange like the New York Stock Exchange (“NYSE”) or the National Association of Securities Dealers Automated Quotations (“NASDAQ”). Instead, retail orders are sent to a wholesale market maker, which is a big firm that trades electronically. The market maker gets the order, fills the order by trading it with its own inventory or matching it with another order, and pays the broker a small amount of money for each share. This PFOF process avoids the old exchange-based trading model and has created a powerful off-exchange system for retail trading. As noted by Battalio and Jennings (2022), this practice allows market makers to internalize retail orders that are typically uninformed, providing a source of liquidity distinct from public lit markets. The money made from PFOF has allowed many brokers to offer "commission-free" trading, which is a powerful way to get new customers and has completely changed the brokerage industry (Battalio and Jennings).

B. Key People Involved in PFOF Transactions (Brokers, Market Makers, Exchanges)

The PFOF system includes several key people, each with a different job and function in the transaction. The main people are the retail brokers, the wholesale market makers, and, in a less important role, the public exchanges.

Retail Brokers: These are the companies that serve regular, non-professional investors. Examples include large companies like Robinhood, Charles Schwab, and Fidelity. For these companies, PFOF has become a very important, and sometimes their biggest, source of money. By not charging trading fees, brokers can attract a huge number of new and active traders. A group of retail traders and their combined orders are the broker's most important asset. The broker's goal is to maximize profits from this asset, which means sending orders to the market maker that offers them the best financial deal. However, the broker's legal duty is to find the "best execution" for their customer. This duty means they must try to get the most favorable terms for the customer. PFOF may seem misaligned with the duties of brokers, as the goal of PFOF is to maximize profit, while the broker is obligated to fulfill the best trade for their customer as ordered.

Wholesale Market Makers: These are financial firms that specialize in filling orders and ensuring sufficient supply to trade. Examples include Citadel Securities and Virtu Financial, among others. Market makers' business model is based on always offering to buy and sell stocks and making money from the small difference between those prices, which is called the spread. PFOF gives them a steady and predictable flow of orders from regular people. The value of these orders is that they are not based on big, secret information that could move prices against them. By filling these orders away from the public exchange, market makers can often give a slightly better price than the best one available on an exchange. Brokers use better pricing to show they are fulfilling their "best execution" duty. The market maker makes a profit by keeping the bid-ask spread and the difference between the price at which they filled the order and the public market price.

Exchanges: Public stock exchanges, like the NYSE and Nasdaq, are the traditional places for stock trading. They are where buyers and sellers meet in a public and open marketplace. The rise of PFOF and off-exchange trading has greatly reduced the number of orders from regular people that go through these public markets. Because of the rise of PFOF, exchanges have had to change how they do business, often by offering their own money and discounts to attract orders. However, large retail exchanges are still at a disadvantage when it comes to getting business from regular people, since their open, auction-style markets often cannot compete with the profitable off-exchange model offered by market makers. This discrepancy in business advantage has caused exchanges to speak out against PFOF, saying that PFOF makes the market less open and fair by taking a large part of the trading activity away from public view. This market fragmentation is harmful because it siphons "uninformed" liquidity away from public venues, widening the bid-ask spreads for all participants and degrading the quality of public price discovery. When exchange liquidity is reduced, the remaining orders on the lit market, often from sophisticated, "informed" institutional players, become more expensive to execute, ultimately increasing the overall cost of capital and making the broader market less efficient.

C. Regulatory Framework and Legal Status of PFOF in the U.S. and Internationally

The rules and legal status of PFOF are very different in various parts of the world, showing a big disagreement on whether the practice is good or bad. Ultimately, the global landscape demonstrates that the existence of PFOF is a deliberate regulatory choice rather than an inherent market necessity.

In the U.S., PFOF is allowed but is heavily regulated. The main laws surrounding PFOF are managed by the Securities and Exchange Commission (“SEC”). Key rules include:

- SEC Rule 606: This rule requires brokerage firms to publicly show how they send their orders, including how much money they get for it. The goal is to be open and allow investors to see if their broker is doing its job of getting them the best trade. While this rule requires them to share information, it does not ban PFOF outright.
- Best Execution: Under rules from the Financial Industry Regulatory Authority (“FINRA”) and the SEC, brokers must provide "best execution" for their customers' orders. Best execution means they must try to get the most favorable terms possible, including price, how fast the trade happens, and how likely the trade is to be completed. The core of the PFOF usage debate in the U.S. is about the best execution concept. Seemingly at odds is whether a broker gets PFOF commission and still honestly fulfills its duty to get the best trade for its customer, especially when another market maker might offer a slightly better price but without the added commission from PFOF. (European Securities and Markets Authority 2015)

In recent years, the PFOF practice has been more heavily scrutinized. In 2022, SEC Chairman Gary Gensler suggested some big changes to make the U.S. market more competitive and open. These proposals did not call for a total ban; instead, they focused on significantly restructuring PFOF by requiring retail orders to be sent to open auctions where multiple firms could compete to fill them, thereby reducing the dominance of private wholesale arrangements while stopping short of getting rid of the practice altogether. The SEC is worried about the problem of brokers having two competing goals and the possibility that a broker might care more about its own profit than its customer's best interest. The result of these ideas is not yet known, but they show that there could be a major change in the rules.

In contrast to the U.S., many other major financial places have either banned PFOF or made it very difficult to do.

- European Union (EU): Under the Markets in Financial Instruments Directive II (“MiFID II”), the EU has effectively banned PFOF. The rule-makers see PFOF as a problem that goes against a broker’s duty to act in the best interest of its customers. MiFID II requires that firms act "honestly, fairly and professionally" when dealing with customers, and PFOF is specifically banned because it is seen as something that could harm this duty. While some small exceptions exist in certain countries, the general feeling in the EU is

that PFOF is not good for protecting investors. (European Securities and Markets Authority 2015)

- United Kingdom (“UK”): The UK’s approach, after leaving the EU, is very similar to the EU’s. The Financial Conduct Authority (“FCA”) has the same worries about the problem of competing goals and has taken steps to ban or severely limit PFOF, seeing it as a practice that could hurt investors.

The big difference between the U.S. and international markets showcases differing opinions on how to handle PFOF. This highlights that the current state of market structure is a product of specific policy choices: the U.S. has historically chosen to allow PFOF under the assumption that transparency and competition can mitigate risks, while the EU and UK have made the policy choice to prioritize the mitigation of conflicts of interest through prohibition. This dichotomy in approach can have drastically different outcomes for retail-based investors and their brokers.

II. Historical Overview of Robinhood Markets, Inc. and PFOF

A. The Origins of Robinhood Markets, Inc.

Robinhood Markets, Inc. was founded in 2013 by Vladimir Tenev and Baiju Bhatt, two Stanford graduates who previously built high-frequency trading platforms for financial institutions. Their unique experience exposed them to the stark inefficiencies and high costs that separated institutional investors from everyday retail traders. At the time, even basic stock trades often carried a commission of \$5 to \$10, making frequent trading and small investments uneconomical for a large segment of the population. Tenev and Bhatt were motivated by the idea of democratizing finance, with the explicit goal of making stock market participation accessible to everyone, regardless of their wealth or experience. Their vision was to build a mobile-first brokerage platform that eliminated commissions and simplified the trading process.

The company's launch in 2014 was met with immediate enthusiasm, attracting a long waitlist of users eager to trade for free. The name "Robinhood" itself was a deliberate nod to its mission of taking from the financial elite (high commissions) and giving to the masses (commission-free trading). The app's user interface was clean, intuitive, and designed to resemble a modern social media platform rather than a complex trading terminal. The app's gamified elements, such as confetti animations for a user's first trade and push notifications about market movements, were a stark contrast to the staid, formal platforms of legacy brokerages. Robinhood’s user base grew rapidly, attracting a younger demographic that was new to investing. This explosive growth demonstrated a clear market demand for a product that broke down the traditional barriers to entry. While legacy brokers historically relied on transparent, per-trade commissions paid by the user, Robinhood’s business model shifted the primary revenue source to PFOF, making the market maker, rather than the investor, the paying client. The core of Robinhood’s disruptive strategy lay not in what it charged for trades, but in how it chose to make

its money; a model that was initially opaque to its burgeoning user base and would later become the source of significant controversy. (Tan 2021)

B. How Robinhood uses PFOF to Make Money

Robinhood's central value proposition to its customers is the offer of "commission-free" trading. This is a powerful marketing tool that has been widely adopted across the industry since Robinhood's introduction. However, the term "commission-free" is a semantic distinction; it does not mean that the service is completely free, although it may be free to the trader. Instead, Robinhood's business model is almost entirely reliant on generating revenue from sources other than direct fees from its customers, and the most significant of these is PFOF. According to its public filings, PFOF has consistently accounted for the vast majority of Robinhood's total revenue, particularly in its earlier years as a private company and through its initial public offering.

The mechanism through which Robinhood generates this revenue is both straightforward and controversial. When a Robinhood customer places an order to buy or sell a stock or option, Robinhood does not route that order to a public exchange. Instead, Robinhood sells the order to a third-party, off-exchange market maker, such as Citadel Securities, Virtu Financial, or G1 Execution Services. These market makers pay Robinhood a per-share or per-contract rebate for the right to execute the trade. The market makers are willing to pay for this order flow because it is "uninformed" and predictable. Unlike institutional trading, which can involve massive, market-moving blocks of shares, retail orders are generally small and do not contain sensitive information. This predictability allows market makers to efficiently manage their risk and profit from a minimal difference between the bid and ask prices. By executing these trades internally, market makers can capture the spread and the PFOF fee, while often providing a fraction of a cent of "price improvement" on the trade, meaning the customer receives a price slightly better than the best available price on a public exchange. This price improvement is key to Robinhood's argument that it is fulfilling its "best execution" obligation to its clients. Critically, however, this marginal price improvement is often small relative to the total bid-ask spread captured by the market maker, leading to concerns that the hidden costs of wider spreads and lost execution quality may actually outweigh the visible benefit of the nominal price gain.

Beyond PFOF, Robinhood generates revenue from other sources as well, including interest on cash balances held by customers (securities lending) and premium services like Robinhood Gold. However, PFOF remains the cornerstone of its financial strategy, making it a powerful case study for how a modern brokerage can monetize its user base without charging explicit trading fees. This model, while lucrative for the company and attractive to its users, has also created a direct and quantifiable conflict of interest that has been at the heart of its most significant challenges.

C. Public Perception, Controversies, and Regulatory Scrutiny Regarding Robinhood's PFOF Practices

Robinhood's meteoric rise was accompanied by a growing public and regulatory debate over its business practices. The core of the controversy stems from the inherent conflict of interest created by PFOF. While Robinhood legally asserts that it seeks "best execution" for its customers, critics argue that the practice of selling orders creates a perverse incentive for the broker to prioritize its own financial gain over the best interests of its clients. The highest-paying market maker may not always be the one that provides the most favorable execution price or fastest service. This creates a moral hazard where the broker benefits from customer trading activity in a way that is not fully transparent to the user.

The most significant event that brought Robinhood's PFOF practices into the public spotlight was the "meme stock" trading frenzy of early 2021, particularly surrounding GameStop ("GME") and AMC Entertainment ("AMC"). As a collective of retail investors organized on social media platforms, they initiated a massive short squeeze against institutional hedge funds. The unprecedented surge in trading volume led to extreme volatility for these stocks. In response, Robinhood, along with several other brokerage firms, suddenly restricted its users from buying shares of these specific stocks, allowing only for the sale of existing positions. The company attributed this decision to the steep rise in capital requirements demanded by clearinghouses to cover potential settlement risks, a requirement directly tied to the high volume of volatile trading. However, this explanation was widely perceived as a betrayal of its "democratizing finance" mission. Critics alleged that the restrictions were imposed to protect the institutional market makers to whom Robinhood sells its order flow, as these firms were suffering massive losses from the short squeeze.

The fallout from the GameStop saga was immense. Robinhood faced public outrage, multiple lawsuits from its users, and congressional hearings. This event galvanized regulatory scrutiny. The Securities and Exchange Commission ("SEC") had already been investigating Robinhood's business practices, and in December 2020, before the GameStop incident, the company paid a \$65MM fine to settle charges that it had misled customers about PFOF. The SEC's complaint stated that Robinhood had failed to disclose that it sold its customers' orders and that this practice resulted in worse execution prices for its users, effectively undermining its claim to offer the most favorable terms. These enforcement actions directly undercut Robinhood's public messaging of "democratizing finance" by revealing that the platform's financial success was built on a lack of transparency that cost its users more than they saved in commissions. This fine, coupled with a later \$70MM fine from the Financial Industry Regulatory Authority ("FINRA") for a series of operational failures, underscored the regulatory body's growing concerns. (U.S. Securities and Exchange Commission 2020, Financial Industry Regulatory Authority 2020)

In the aftermath of the GameStop controversy, SEC Chairman Gary Gensler openly questioned the value of PFOF to the average investor. Gensler has floated proposals for a major overhaul of market structure, including requiring off-exchange trading to go through auctions or even a potential ban on PFOF. These proposals include the Order Competition Rule and Regulation Best Execution. The first rule would require certain individual investor orders to be

exposed to a "qualified auction" to promote competition for retail order flow and ensure better prices. The second rule would require broker-dealers to have written policies and procedures to achieve best execution for customer orders. While these proposals are not yet finalized, they represent a fundamental challenge to Robinhood's business model and the entire ecosystem of commission-free trading. Robinhood's journey demonstrates that while PFOF has been an engine for democratizing access to financial markets, it has also created profound ethical and regulatory questions that will continue to shape the future of the industry.

III. How PFOF is used in the brokerage industry

A. The Mechanics of PFOF in Modern Brokerage Operations

When an investor uses a free trading app to buy or sell a stock, a complex and quick process typically occurs to execute a stock trade. If a brokerage firm is utilizing PFOF, the trade process usually proceeds as follows:

1. **The Investor Places an Order:** The process begins when a regular person decides to buy or sell a security, like 100 shares of a company. They enter the order into their brokerage's app or website.
2. **Order Routing:** Instead of sending this order directly to a public stock exchange (like the NYSE), the brokerage firm sends it to a wholesale market maker. A market maker is a big financial firm that trades a lot of stocks very quickly. The brokerage has a deal with this market maker to send all of its customers' orders to them.
3. **The PFOF Payment:** When the market maker gets the order, it pays the brokerage firm a small amount of money for it. This payment is the "Payment for Order Flow." The fee is usually a fraction of a cent per share, but when you add up the millions of orders that a company like Robinhood handles every day, it becomes a very large source of income.
4. **Order Internalization and Execution:** The market maker then takes the order and fills it from its own supply of stocks (this process is "internalization"). The market maker acts as the buyer for a sell order and the seller for a buy order. They achieve this by offering the investor a price that is typically slightly better than the best available on a public exchange. This tiny improvement is called "price improvement." The market maker makes money from this by keeping the small difference between the price at which they bought the stock and the price they sold it for, as well as by collecting the PFOF fee from the broker.

This system is different from direct routing or commission-based. In the old way, a broker would send an order to an exchange, where it would be matched with another order from another investor in an open and transparent auction. In the PFOF model, the trade happens in private, away from the public market. Market makers like this system because the orders from regular people are usually "uninformed," meaning they are not based on special knowledge that

could hurt the market maker. This makes it easier for the market maker to predict prices and make money from a large number of small trades. The value of this kind of predictable order flow is why they are willing to pay so much to get it.

B. Major Industry Players and Their Approaches to PFOF

Different market makers and brokers utilize PFOF in different ways. The use of PFOF depends on how a brokerage chooses to make money. The major players can be divided into a few groups:

Commission-Free Platforms (e.g., Robinhood): These companies are the most famous examples of the PFOF model. For a company like Robinhood, PFOF isn't just one source of money; it's the main reason their business works. They offer trading for free, and in return, they use PFOF to make almost all of their money. This model allows them to attract millions of new and young investors who might not have been able to afford trading fees. For these companies, the ability to sell their customers' orders to market makers is their most important asset. Without PFOF, their business model would not be able to work as it does today. This singular reliance makes them the most vulnerable to regulatory shifts, as a ban or severe restriction on PFOF would essentially dismantle their primary revenue stream.

Hybrid Brokers (e.g., Charles Schwab, Fidelity): These are older, more traditional brokerage firms that have been around for a long time. In the past, they made their money from trading fees. As free trading gained popularity due to companies like Robinhood, these brokers had to adapt their business to stay competitive. They eliminated their own trading fees but did not fully transition to the PFOF model. Instead, they use a mix of PFOF and other ways to make money, such as interest on customer cash and management fees for investment accounts. This means they are not as dependent on PFOF as Robinhood is, which gives them more flexibility. They can still claim that their main goal is to get the best deal for their customers because PFOF is not the only way they make money. Because their income is diversified, they are better positioned to weather regulatory changes by pivoting back toward service fees or cash-sweep programs.

Full-Service Brokers (e.g., some smaller firms): These are brokerage firms that offer much more than just a place to trade stocks. They often provide financial advice and other services to their customers, generating revenue from high advisory fees. For these firms, PFOF is not a big part of their business, and some may not use it at all. Their customers are often wealthier and are willing to pay for expert advice, so the brokerage does not need to sell their orders to a market maker to make a profit. In this model, the relationship between the client and the broker is more personal, and there is less concern about the kind of conflicts of interest that PFOF can cause. These firms are effectively insulated from PFOF regulation, as their value proposition is built on expertise rather than low-cost execution.

C. Comparative Analysis of PFOF Practices Between Traditional Brokers and Commission-Free Platforms

The way PFOF is used by different types of brokerage firms is a key part of the modern debate over its use. A side-by-side comparison of the practices of new, free platforms and older, more traditional brokers helps to show the main differences.

Whether or not a broker or financial institution decides to utilize PFOF largely depends on whether such an arrangement fits within its business model and contributes to the profitability of the company. For free platforms like Robinhood, PFOF is the main reason they exist. PFOF provides a direct incentive for the broker to focus on its own profits. The more orders it sells, and the better the price it gets for those orders from market makers, the more money it makes. The profitability incentives from PFOF can sometimes go against the goal of getting the best price for the customer. In contrast, traditional and hybrid brokers have many different ways to make money and do not have to rely on PFOF. Such a lack of reliance on PFOF reduces the financial reason to get money from market makers and, in theory, makes it easier for them to prioritize their customers' best interests.

Government agencies like the SEC require all brokers to share information about their PFOF practices. However, how this information is shared and understood is different. New platforms often market themselves as being "free," which can make it hard for customers to understand how the company actually makes money. The details of the PFOF deals are often hidden in complex reports, and many customers are either unaware of or indifferent to finding them. Traditional brokers, on the other hand, have often been more open about their costs and fees for many years, so the move to a more PFOF-based model might be clearer to their customers. However, the problem of making the process easy for the average person to understand is a problem for the whole industry.

The wide use of PFOF by new, free platforms has completely changed the stock market. Because so many orders from regular people are now being sent to private market makers, a large amount of the trading in the U.S. now happens away from public exchanges. The use of PFOF has led to a more divided market, where it is harder to see all the buying and selling that is happening, which is a big concern for regulators, who worry that a fragmented market could be less fair and harder to watch. The structural vulnerability of purely commission-free firms highlights why current SEC proposals targeting order competition are so contentious; while a diversified hybrid broker might simply adjust its fee structure, a PFOF-dependent platform faces an existential threat to its core operating logic.

While PFOF has made stock trading free for millions of people, its use is very different across the brokerage industry. It is the very core of the business model for new platforms, but it is just one part of the business for older firms. The use of PFOF will continue to be a source of debate, as the desire to make investing accessible to everyone is weighed against the potential conflicts of interest and the changes it has caused to how the financial markets work.

IV. Is PFOF beneficial to the retail trader?

A. Advantages of PFOF for Retail Investors

One of the biggest benefits that retail traders and investors gain from PFOF is that no trading fees are charged to them. In the past, a single stock trade would cost anywhere between \$5 to \$10 per trade, and for a new investor with a small amount of money, these fees could be quite costly. A \$10 fee on a \$100 trade meant an instant 10% loss. In order for the position to become profitable, the stock value would have to appreciate to make up the difference. Brokerages that use PFOF have made it much easier for people to get started by not charging any fees for stock trades. This has allowed new groups of investors to join the market by letting them regularly invest small amounts of money and build portfolios over time. Getting rid of fees also lets investors trade more often without worrying about the cost, which can be helpful for those who want to buy and sell small parts of different stocks. However, this increased accessibility introduces complex behavioral consequences: the elimination of friction may encourage overtrading, where investors transact more frequently than is optimal, potentially leading to lower long-term returns and complicating the overall welfare analysis of zero-cost trading.

Another key advantage that brokerages often talk about is “price improvement”. When a brokerage sells a customer's order to a market maker, the market maker has a reason to provide a price that is slightly better than the best price on a public exchange. For example, if the best price to buy a share of a stock is \$20.00, the market maker might fill the order at \$19.998. These tiny savings are the “price improvement.” While this amount is very small for one trade, brokerages argue that it can add up to a lot of money for investors who trade often. In this view, PFOF actually helps the customer by giving them a better price than they might have gotten on a public exchange, all while avoiding the fee they would have paid. This practice is seen as a way for the broker to follow its legal duty of “best execution” for its clients, making sure the client gets the most favorable terms for their trade.

In contrast, industry groups and many of the brokerages that use PFOF defend the practice by pointing to its overall effect. They argue that focusing on individual, tiny costs misses the bigger picture. By allowing millions of people to invest who would have been unable to afford it otherwise, PFOF has opened up finance and created a more inclusive market. From this point of view, the benefits of getting more people to invest are more important than the possible problems of a less-than-perfect execution price.

B. Disadvantages and Potential Conflicts of Interest for Retail Investors

Despite the claimed benefits, the PFOF model also has significant problems and a clear chance for conflicts of interest. At its core, PFOF creates a classic principal-agent problem: the investor (the principal) relies on the broker (the agent) to act in their best interest, but the broker’s compensation is tied to the market maker rather than the client. The most serious concern is that a brokerage's financial reason to make money from PFOF may not match its duty to provide the best service to its clients. This is an explicit conflict of interest because the broker is incentivized to route orders based on the highest rebate offered by a market maker, which may come at the expense of the investor receiving a superior execution price. The most profitable deal for the brokerage might not be the best deal for the investor. For example, a market maker who

pays the most in PFOF fees might not offer the best prices. This creates a situation where the broker is paid to send its customers' orders to a specific firm, which could lead to a less favorable result for the customer. In addition, a brokerage may not disclose to its clients who is acting as the brokerage's market maker, and the clients may not have the ability to track how their trades are being executed. The opaqueness of the entire PFOF process on the client side means that retail trades may not know if their broker is working for their best interests or for the interests of the firm that is paying them.

Another major problem with PFOF is the lack of transparency about the true cost of a trade. While fees have been removed, the investor still "pays" in a different way. The market maker makes money by taking a small part of the "bid-ask spread, which is the difference between the highest price a buyer is willing to pay and the lowest price a seller is willing to accept. When a market maker fills an order from their own supply of stocks, they are capturing this spread, and the PFOF fee is basically a way to share that profit with the brokerage. The investor is often unaware that this is happening and has no way to know if the price improvement they received truly makes up for what the market maker is gaining from the spread. This hidden cost can add up, potentially exceeding the fees charged in the past.

The PFOF model also creates a problem that may raise regulatory concerns because the act of purchasing order flow from brokerages gives an informational advantage that undermines concepts of market fairness and openness. Market makers who buy order flow from brokerages get to see the orders before anyone else. They know that a stream of small, retail orders is coming. Even if market makers are not allowed to use this information to trade for themselves before filling the customer's order, they can use this information to adjust their prices. The receipt of trading information and order flow gives an advantage over other traders in the market, creating an unfair playing field. From the Investopedia article "SEC Considers Banning Payment for Order Flow": "The U.S. Securities and Exchange Commission (SEC) is considering a full ban on the payment for order flow (PFOF). The reason is that this practice creates "an inherent conflict of interest," according to SEC Chairman Gary Gensler, in a recent interview with Barron's." Given the PFOF practice, regulators are concerned about equal access and fair trading that may not be prevalent if market makers can have free rein over the pricing and execution of a large amount of retail investor trades.

One of the most important findings came from the Securities and Exchange Commission ("SEC") in its December 2020 settlement with Robinhood. In the Press Release: "SEC Charges Robinhood Financial With Misleading Customers About Revenue Sources and Failing to Satisfy Duty of Best Execution", the SEC accused the company of not being open about its PFOF practices. (U.S. Securities and Exchange Commission 2020) It found that, despite claiming to provide "superior" execution, Robinhood often sent its customers' orders to firms that gave worse prices than other market makers. The SEC found that this caused customers to lose more than they saved from not paying fees. This finding directly challenged the brokerage's main claim that PFOF was a good thing for its customers and showed the serious conflicts of interest that the model creates.

Conclusion

On its face, PFOF seemingly helps retail traders by offering them a way to trade without any “trading fees.” However, as Warren Buffett has said, “There is no free lunch”. This quote embodies his belief that you can't get something for nothing, and any desired or valuable outcome comes with a cost, whether in money, effort, or sacrifice. On one hand, PFOF has, without a doubt, opened up the world of investing to a wider group of people, getting rid of the direct cost of trading and providing a smoother experience. On the other hand, it has created a new kind of "cost" that is less transparent and involves significant conflicts of interest. This implicit cost is realized not through a visible commission but through reduced price improvement, which, when aggregated across millions of trades, transfers billions of dollars annually from retail investors to market makers.

Moreover, when so many trades bypass major exchanges, it breaks up the market, making it harder to find the true best price for a stock and granting dominant market makers an unfair informational advantage. The findings from regulators and experts suggest that these hidden costs can be substantial, and that the financial interests of the broker and the market maker may not always align with the best interests of the investor. In addition, the incentives of the main purveyors of PFOF, such as brokerages and the related market makers, may not be aligned with the greater good of the market as seen through the GameStop Controversy.

While Payment for Order Flow has been a powerful engine for expanding access to financial markets, its systemic conflicts of interest and hidden costs constitute a clear failure of market fairness, inflicting far more damage on the core principles of a free and open market than any benefit it provides.

Works Cited

- Admati, Anat R., and Paul Pfleiderer. "Selling and Trading on Information in Financial Markets." *The American Economic Review*, vol. 78, no. 2, 1988, pp. 96-103.
https://www.jstor.org/stable/1818105?searchText=&searchUri=&ab_segments=&searchKey=&refreqid=fastly-default%3Acaa34eb8fa599c30c8c23e4a48ab757b&initiator=recommender&seq=
- Barber, Brad M., et al. "Attention Induced Trading and Returns: Evidence from Robinhood Users." *Journal of Finance*, vol. 77, no. 6, 2022, pp. 3141-3190.
<https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.13183>
- CFA Institute. "Payment for Order Flow in the United Kingdom." CFA Institute Research Foundation, 2016.
<https://rpc.cfainstitute.org/policy/positions/payment-for-order-flow-in-the-united-kingdom>
- CNBC. "Payment for Order Flow: The Controversial Wall Street Practice to Draw Scrutiny at Robinhood Hearing." CNBC Finance, 2021.
<https://www.cnbc.com/2021/02/18/payment-for-order-flow-the-controversial-wall-street-practice-to-draw-scrutiny-at-robinhood-hearing.html>
- CNBC. "Virtu Financial CEO Says Pay for Order Flow Benefits Retail Investors." CNBC Video Interview, Sept. 8, 2021.
<https://www.cnbc.com/video/2021/09/08/virtu-financial-ceo-says-pay-for-order-flow-benefits-retail-investors.html>
- Eaton, Green, et al. "Retail Trader Sophistication and Stock Market Quality: Evidence from Brokerage Outages." *Journal of Financial Economics (JFE)*, Vol. 146, 2022, pp. 502-528
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3776874
- European Securities and Markets Authority (ESMA). "Guidelines on Complex Debt Instruments and Structured Deposits." 2015.
https://www.esma.europa.eu/sites/default/files/library/2015/11/08_735.pdf
- European Securities and Markets Authority (ESMA). "Public Statement: Payment for Order Flow (PFOF)." ESMA35-43-2749, 2021.
https://www.esma.europa.eu/sites/default/files/library/esma35-43-2749_esma_public_statement_pfof_and_zero-commission_brokers.pdf
- Financial Industry Regulatory Authority (FINRA). "FINRA Reminds Member Firms of Their Obligations Regarding Best Execution." Regulatory Notice 21-23, 2021.
<https://www.finra.org/rules-guidance/notices/21-23>
- Gensler, Gary. "Prepared Remarks at the Global Exchange and Fintech Conference." U.S. Securities and Exchange Commission, 2021.
<https://www.sec.gov/newsroom/speeches-statements/gensler-global-exchange-fintech-2021-06-09>
- Tan, Kuo. "Democratizing finance with Robinhood: Financial infrastructure, interface design and platform capitalism" Volume 53, Issue 8 of *Environment and Planning A: Economy and*

Space, 2021.
https://www.researchgate.net/publication/354570809_Democratizing_finance_with_Robinhood_Financial_infrastructure_interface_design_and_platform_capitalism

Robinhood Markets, Inc. "Form S-1 Registration Statement." U.S. Securities and Exchange Commission, 2021.
<https://www.sec.gov/Archives/edgar/data/1783879/000178387924000083/hood-20240320.htm>

U.S. House of Representatives Committee on Financial Services. "Game Stopped? Who Wins and Loses When Short Sellers, Social Media, and Retail Investors Collide." 117th Congress, 2021.
<https://www.congress.gov/event/117th-congress/house-event/112590>
<https://www.govinfo.gov/content/pkg/CHRG-117hhrg44837/pdf/CHRG-117hhrg44837.pdf>

U.S. Securities and Exchange Commission. "In the Matter of Robinhood Financial, LLC: Order Instituting Administrative and Cease-and-Desist Proceedings." Securities Act Release No. 10906, 2020.
<https://www.sec.gov/files/litigation/admin/2020/33-10906.pdf>

Wharton Initiative on Financial Policy and Regulation (WIFPR). "Payment for Order Flow." Wharton University of Pennsylvania, 2023.
<https://wifpr.wharton.upenn.edu/wp-content/uploads/2023/09/Payment-for-Order-Flow-final.pdf>

The Promise of Fusion Energy and Challenges after Breakeven with Commercial Reactors

By Vatshal Gaur

Introduction

Fusion energy is a promising future source of energy, as it is safe and potentially nearly infinite. Unlike fossil fuels, which are the main source of the world's energy today, fusion produces no carbon dioxide, thus providing support to reduce the effects of climate change. Additionally, applications of fusion energy extend beyond generating electricity. Fusion can provide high-temperature heat for industrial processes such as steel making, cement production, or chemical manufacturing. Another application is to capture carbon from the air. Once fusion becomes fully viable at a large scale, it can work alongside solar and wind, keeping grids stable and reliable even when the sun's or wind's energy output drops [1]. In 2022, researchers at the National Ignition Facility in California achieved "ignition," a milestone for fusion energy in which they produced more energy from the fusion reaction than the energy delivered to target [2][3]. Although commercial use of fusion energy is still years away, breakthroughs like "ignition" demonstrate how fusion can reshape the world's energy future.

Our World in Data (2024) reports that fossil fuels cause severe health impacts due to air pollution created by them. For example, coal-fired power plants provide approximately 40% of the world's electricity, but in return produce harmful byproducts such as bottom ash, a toxic byproduct containing heavy metals that can contaminate air, soil, and water [4]. Contrary to fossil fuels, fusion creates very small amounts of short-lived radioactive waste and thus far less long-term impact on the environment. Fusion is a great potential solution to provide continuous clean energy that strongly aligns with global efforts to achieve net-zero emissions and build a more sustainable future.

Background Theory

Nuclear fusion is defined as a process in which a substantial amount of energy is released as a result of the combination of two light atomic nuclei, most likely isotopes of hydrogen, to form a heavier element such as helium [1]. To achieve this process, extremely high temperatures and pressures are required for nuclear collision and ultimately fusion. Fusion reactions take place in a special state of matter called plasma, which is a hot, charged gas made of positive ions and free-moving electrons with unique, distinct properties from solids, liquids, and gases, and provides a vital environment for fusion.

The sun and other stars are powered by nuclear fusion. Inside the sun, nuclei collide at extremely high temperatures, around 10 million degrees Celsius, to fuse together [1]. These high temperatures provide nuclear energy to overcome their mutual electrical repulsion. The Sun is made mostly of hydrogen nuclei that combine to form helium nuclei. When hydrogen fuses into helium, a small amount of mass is lost and is converted to energy according to Einstein's equation $E = mc^2$. This energy is released as light and heat, which travels from the sun's core to its surface and out into space. This fusion process happens continuously, keeping the sun powered.

Similarly, on Earth, fusion fuel is accessible. The hydrogen isotopes required for fusion reactions are Deuterium, which can be extracted from seawater, and Tritium, which can be produced from lithium. At present, scientists are pursuing two methods for achieving nuclear fusion: inertial confinement fusion and magnetic confinement fusion. Both methods aim to heat and confine hydrogen isotopes long enough to overcome electrostatic repulsion and fuse, releasing energy in the process.

Inertial Confinement Fusion (ICF)

Inertial confinement fusion The pellet is surrounded by a light material that vaporizes outwards when heated by the lasers. And when it does, there is a net reaction inward, driving an implosion. This is similar to a spherical rocket. By ejecting exhaust outwards, it drives the rocket in the opposite direction. In this case, the vaporized material on the outside of the pellet pushes the hydrogen isotopes in toward the center. The lasers must be applied accurately to get a symmetrical shock wave moving toward the center of the hydrogen mixture – creating the temperature and density needed to start the fusion reaction. NIF ignition events use 192 laser beams to create this implosion and cause the isotopes to fuse.

Magnetic Confinement Fusion (MCF)

Magnetic confinement fusion uses intense magnetic fields to confine plasma. Scientists use special devices to trap plasma in a closed magnetic loop as fusion plasma reaches very high temperatures of hundreds of millions of degrees Celsius, so no solid material can contain them directly. The magnetic field prevents the hot plasma from touching the reactors' walls. The most promising fusion reactors of that kind are tokamaks and stellarators. The tokamak is a doughnut-shaped vacuum chamber where hydrogen gas under intense heat and pressure turns into plasma that allows nuclei to fuse and release energy. Magnetic coils surrounding the chamber create a strong magnetic field to confine plasma and prevent it from touching the vessel walls. Magnetic systems and electric current ionize the gas, forming plasma.

On the other hand, a stellarator also confines plasma in a toroidal shape but uses external 3D magnetic coils that are twisted and shaped precisely to generate a magnetic field. Both reactors have advantages and disadvantages. Tokamaks are currently more common because of their easier design and the capability to heat plasma efficiently to very high temperatures, but they pose a challenge in keeping plasma stable. Stellarator configurations can keep plasma stable but are challenging to build, leading to being the preferred solution in the future.

What Has Been Done Till Now

Rapid advancement in material science, magnet technology, and reactor design have brought the vision of fusion energy closer to reality.

Advances in Inertial Confinement Fusion (ICF)

The National Ignition Facility (NIF) in the United States achieved a historic milestone in December 2022, when scientists produced more energy from fusion than the energy delivered by

the lasers to the target. This event, “ignition,” proved that controlled nuclear fusion is physically achievable. What made it possible was an improved laser symmetry of 192 laser beams with unprecedented precision to minimize asymmetries, which are fired into a centimeter-sized hollow cylinder called a hohlraum. Since then, NIF has conducted several successful follow-up experiments, including the April 2025 experiment, which delivered 8.6 megajoules (MJ) of fusion energy with a target gain of 4.13, marking one of the highest yields ever recorded.

In 2020, Osaka University's Institute of Laser Engineering under FIREX (Fast Ignition Realization Experiment), produced the same level of energy with less than 1/10 of the energy inputted than previously. This technique, known as 'fast ignition,' needs less total energy and could lead to making laser fusion reactors smaller, cheaper, and more efficient [5].

The Z machine at Sandia National Laboratories is the world's largest X-ray generator, using powerful electrical pulses to create Z-pinch plasma that produces intense X-rays capable of compressing fusion fuel. In 2006, it achieved a temperature above 2 billion °C, far exceeding the threshold for hydrogen fusion and even allowing theoretical fusion with heavier elements like lithium or boron [6].

Advances in Magnetic Confinement Fusion (MCF)

The International Thermonuclear Experimental Reactor project (ITER), an international collaboration involving 35 countries (including the EU, USA, Russia, China, Japan, India, and South Korea), is under construction in southern France and will be the world's largest tokamak. Once operational, ITER aims to produce 10 times more energy (500 MW) than the power used to heat its plasma, making it the world's most ambitious step toward a sustainable fusion power plant [7].

In a groundbreaking achievement in pursuit of continuous and self-sustaining fusion reactions, in February 2025, the WEST tokamak reactor in France set a new record to maintain hot plasma for 22 minutes while reaching a temperature of 50 million °C. The previous plasma duration record of 1066 seconds was achieved in Jan 2025 by China's Experimental Advanced Superconducting Tokamak (EAST), often referred to as "artificial sun," at 70 million °C. Previously, South Korea's KSTAR reactor successfully maintained 100 million-degree plasma for 30 seconds in December 2021 [8].

In the United Kingdom, the now decommissioned Joint European Torus (JET) was a leader in magnetic confinement fusion. In 2024, JET achieved a record 69 megajoules of fusion energy, demonstrating the effectiveness of deuterium-tritium plasma in large-scale experiments and paving the way for the next-generation ITER reactor.

Wendelstein 7-X, Germany, the leading fusion device of the stellarator type, generated long-lasting plasma pulses for 43 seconds with record confinement and stability. This is a new world record for triple product in a long plasma discharge [9].

Researchers in the U.S. Department of Energy's (DOE) Princeton Plasma Physics Laboratory (PPPL) combined decades of expertise in engineering, computation, and theoretical physics to design a new type of stellarator, MUSE, using permanent magnets instead of a

superconducting electromagnet system like W-7X. This will pave the way for simpler and less expensive designs [10].

Together, these global efforts represent extraordinary scientific collaboration. Each milestone from NIF's ignition to JET's high-yield runs and ITER's construction brings the world closer to realizing the long-term dream of clean, limitless, and sustainable fusion energy.

What Needs to Be Done

Achieving early-stage fusion power plants by the 2030s requires emphasis on fusion materials science, plasma-facing components, tritium-breeding blanket technology, and the tritium fuel cycle. Several key experimental facilities are recommended. The Fusion Prototypic Neutron Source (FPNS) will provide unique material irradiation capabilities, and the Material Plasma Exposure eXperiment (MPEX) and high-heat-flux testing experiments will enable solutions for the plasma-facing materials. Blanket research and associated integration blanket facilities will provide the scientific understanding and basis to qualify fusion fuel breeding systems for future power plants [11].

Addressing the technology and science drivers will require continuing investment in the design, construction, and operation of facilities that provide important new capabilities. To sustain this progress, consistent funding and program-level resources must be established alongside new small and mid-sized scale research facilities. Programs like INFUSE (Innovation Network for Fusion Energy) to strengthen collaboration between private industries and national laboratories need to expand to agencies such as the National Science Foundation (NSF), ARPA-E, and the National Nuclear Security Administration (NNSA). These partnerships between the public and private sectors will help reduce costs, accelerate innovation, and share risks. In parallel, policy changes should be developed and implemented that improve diversity, equity, and inclusion within the research community and allow discipline-specific workforce development. Universities and other public institutions require increased funding to advance foundational fusion research while educating and training a new workforce. Funds for these purposes are already available via the CHIPS and Science Act, which authorizes support for non-semiconductor technologies, including advanced energy technologies. The Tech hub program under the CHIPS and Science Act was approved for \$10B over 5 years. So far, only \$1B has been appropriated. Congress can support the first key fusion challenge by appropriating available CHIPS and Science funds, ensuring that the U.S. remains a leader in the race to bring fusion energy to the grid [11].

The other challenge towards commercialisation of fusion energy, once a successful proof of concept is developed, is engineering a cost-competitive fusion power plant. The Levelized Cost of Energy Comparison LCOE of utility-scale solar PV farms ranges from \$0.029 to \$0.092 per kWh, and that of natural gas power plants is only slightly higher(\$0.045 to \$ 0.108per per kWh), but for Nuclear Energy is way higher as of now \$0.141 to \$0.221 per kWh. In order to be competitive, fusion energy must aim below \$0.06-\$0.08 kWh with these mature renewables and efficient gas plants [12].

Conclusion

Although fusion is not ready for commercialisation yet, it has the power to reshape global energy systems. Its energy density and power far exceed that of any other energy source, making it a strong candidate to become the “final energy source” once optimized for grid use. By investing in the fusion research ecosystem today, through supportive policies and sustained investment, the world can accelerate the transition towards a clean, limitless, and net-zero emission energy. Strong leadership, strategic investment, and international collaboration will be key to success.

Works Cited

- 1.Barbarino, M. (2023, August 3). What is Nuclear Fusion? International Atomic Energy Agency. <https://www.iaea.org/newscenter/news/what-is-nuclear-fusion>
- 2.U.S. Energy Information Administration. (2023, August 21). Nuclear Explained - U.S. Energy Information Administration (EIA). Eia.gov; U.S. Energy Information Administration. <https://www.eia.gov/energyexplained/nuclear/>
- 3.Ritchie, H. (2020, February 10). What are the safest and cleanest sources of energy? Our World in Data. <https://ourworldindata.org/safest-sources-of-energy>
- 4.Ritchie, H., & Rosado, P. (2025, March 31). Air pollution kills millions every year — where does it come from? Our World in Data. <https://ourworldindata.org/air-pollution-sources>
- 5.Ryosuke, K. (2020). Fast Ignition Laser Fusion Energy Research in Japan. Available at: <https://conferences.iaea.org/event/214/contributions/17041/>
- 6.Science X (2006). Z machine exceeds two billion degrees Kelvin: hotter than the interiors of stars. Available at: https://phys.org/news/2006-03-machine-billion-degrees-kelvin-hotter.html#google_vignette
- 7.ITER (2023). In a Few Lines. Available at: <https://www.iter.org/few-lines>
- 8.K.Dulon (2025). Achievements at EAST and WEST. Available at: <https://www.iter.org/node/20687/achievements-east-and-west>
- 9.Admin (2024). JET beats its own record. Available at: <https://www.iter.org/node/20687/jet-beats-its-own-record>
- 10.Princeton Plasma Physics Laboratory (2024). A return to roots: PPPL builds its first stellarator in decades and opens the door for research into new plasma physics. Available at: <https://www.pppl.gov/news/2024/return-roots-pppl-builds-its-first-stellarator-decades-and-opens-door-research-new-plasma>.
- 11.Us Fusion and Plasmas (2025). Powering the Future Fusion & Plasmas. Available at: <https://usfusionandplasmas.org/report/executive-summary/>
- 12.Lazard (2024). Lazard’s Levelized Cost of Energy Analysis. Available at: https://www.lazard.com/media/xemfey0k/lazards-lcoeplus-june-2024-_vf.pdf

The Application of Stem Cells as a Treatment for Glioblastoma

By Baek, H.J., Pearson, A.N.

Abstract

Glioblastoma multiforme (GBM) is an aggressive, therapy-resistant brain tumor with rapid progression, high recurrence, and poor survival rates. Current treatments are restricted by the blood–brain barrier (BBB), which effectively blocks the passage of many chemotherapeutic and biologic drugs. Tumor heterogeneity enables resistant cell populations to survive and repopulate. These combined barriers showcase the urgent need for therapies that can improve delivery, counter immune evasion, and effectively target cancer stem cells (CSCs). Stem cell-based therapies offer a possible solution for treating GBM because of their tumor-homing ability, their ability to cross the BBB, and their versatility when genetically engineered to carry therapeutic cargo. Stem cells can transport cytotoxic agents or gene therapies into tumors, secrete cytokines to remodel the tumor microenvironment, or be designed to interfere with CSC-driven mechanisms of recurrence. In this review, we examine three applications of stem cells in GBM therapy: delivery platforms using exosomes or engineered mesenchymal stem cells, immunomodulatory approaches with cytokine-secreting stem cells, and CSC-driven strategies that disrupt tumor renewal. Together, these methods show strong potential in overcoming barriers to GBM treatment. Although clinical translation of stem cell-based GBM therapies remains limited, they hold unique potential to transform treatment by improving delivery, enhancing anti-tumor immunity, and addressing recurrence at its root. Future work should refine targeting specificity and establish the long-term safety of stem cell therapies in glioblastoma.

Keywords

Stem cell therapy, glioblastoma, blood-brain-barrier, immunomodulation, cancer stem cells.

Introduction

Glioblastoma (GBM) is the most common primary malignant brain tumor, affecting roughly 17,000 people in the United States annually (Omuro, 2013). GBM is characterized by rapid growth, invasive spread into surrounding brain tissue, and frequent genetic alterations such as isocitrate dehydrogenase (IDH) mutations, MGMT promoter methylation, and EGFR amplification, challenging the early diagnosis (Aldape, 2015). Conventional treatments, including surgical resection, radiotherapy, and temozolomide chemotherapy, offer limited survival, owing to tumor recurrence due to therapy-resistant cancer cell populations (Wirsching, 2016). A major clinical obstacle in GBM management is the restriction offered by the blood–brain barrier (BBB), a network of tightly connected endothelial cells that selectively limits the passage of molecules from the bloodstream into the brain (Alahmari, 2021). While the BBB protects normal brain tissue, it blocks most drugs because therapeutic agents are limited by their

molecular weight and/or hydrophilicity to cross the junctions, making it challenging to deliver therapies that target GBM (Wu, 2021).

Stem cells emerged as a potential solution owing to their potential to migrate across the BBB, inherent ability to home at tumor sites, chemokine signaling, and inflammatory responses (Alahmari, 2021; Nowak, 2021; Wu, 2021). Additionally, their ability for self-renewal and differentiation into multiple lineages promotes therapeutic versatility in GBM management (Biswas, 2007; Nowak, 2021; Wu, 2021). Stem cells are undifferentiated cells with unique properties of self-renewal and differentiation to multiple lineages, drawing significant attention for regenerative medicine and cancer therapy.

Mesenchymal stem cells (MSCs), a kind of adult multipotent stem cell found in bone marrow, adipose tissue, and umbilical cord blood, have a strong potential for GBM therapy and drug targeting (Nowak, 2021). Embryonic stem cells (ESCs), derived from the inner cell mass of blastocyst-stage embryos and capable of differentiating into all cell types of the body, are potential targets. However, ethical concerns and practical limitations in harvesting hurdles their application in GBM applications (Biswas, 2007). Interestingly, the growing body of evidence signifies the strong potential of stem cells in GBM management. On this background, this review comprehends the stem cell-based treatment strategies and challenges in depleting the BBB for GBM management.

Stem Cell-Based Drug and Gene Delivery Methods

Traditional treatments for glioblastoma, including chemotherapy and radiation, greatly improve patient survival. However, the complete cure of GBM remains unattained owing to the BBB. Stem cells are superior in GBM research due to their programmability, access to tumor site, BBB permeability, and tunability as drug carriers (Wu, 2021). Importantly, recent findings reflected versatile stem cell populations, including ESCs and MSCs possessing tumor-tropic properties (Nowak, 2021). Notably, the biological origin of these stem cells influences their tumor-targeting capacity and functional integration within the tumor microenvironment, making it an essential consideration in therapeutic design (Nowak, 2021).

Zhu et al. demonstrated that engineering embryonic stem cell-derived exosomes (ESC-exos) improved glioblastoma (GBM) drug bioavailability by permeabilizing the BBB (Wu, 2021). BBB challenges the bioavailability of the commonly used chemotherapeutic agent paclitaxel (PTX). Engineering ESC-exos with the tumor-targeting peptide c(RGDyK) PTX significantly enhanced the sustainability of the drug in GBM tissue. This approach significantly improved therapeutic efficacy (Zhu, 2019). ESC-exos were functionalized and drug-loaded using incubation-based techniques. *In vitro* experiments using U87 and U251 glioblastoma cell lines demonstrated that the modified exosomes showed superior cellular uptake and significantly reduced tumor cell proliferation compared to free PTX or unmodified exosomes (Zhu, 2019). Also, apoptosis induction in GBM cells was confirmed in the exosome-PTX-treated cells. *In vivo* biodistribution and therapeutic performance using subcutaneous and intracranial GBM xenograft mouse models revealed greater tumor-specific accumulation of cRGD-Exo-PTX, confirming

successful delivery across the BBB (Zhu, 2019). In the intracranial model, treatment with cRGD-Exo-PTX significantly extended median survival time, demonstrating the superior effects of functionalized ESC-exos as drug carriers, validating their therapeutic potential for GBM.

MSCs are known for their tumor-tropic properties (Nowak, 2021), which enable them to migrate toward tumor sites through chemokine signaling pathways. CXCR4/SDF-1 axis-activated migratory behavior promotes them as candidates for delivering therapeutics to glioblastoma (GBM). The lack of a consistent and renewable source of MSCs limits clinical translation of MSC-based therapies. Addressing this, Bak et al demonstrated MSC-like cells derived from human embryonic stem cells (hESCs) as a scalable delivery platform for prodrug gene therapy in GBM (Bak, 2011). The study programmed hESCs to express the herpes simplex virus thymidine kinase (HSV-TK) gene, enabling to convert the prodrug ganciclovir (GCV) into a toxic compound selectively within the tumor microenvironment of immunocompromised mice with intracranial glioma xenografts. The outcome revealed that tumor growth was significantly reduced and survival was extended, directly supporting the cell delivery system for therapeutic benefit.

Seminal research investigates stem cell-based strategies for delivering therapeutics to glioblastoma (GBM) using different delivery vehicles. Zhu et al designed a nanoparticle-based approach by modifying embryonic stem cell-derived exosomes (ESC-exos) with the targeting peptide c(RGDyK) and loading them with paclitaxel (PTX), enabling drug delivery across the blood-brain barrier (BBB). Contrastingly, Bak et al employed mesenchymal stem cell-like (MSC-like) cells derived from human embryonic stem cells, genetically engineered to express the herpes simplex virus thymidine kinase (HSV-TK) gene, to locally activate the prodrug ganciclovir (GCV) within the tumor microenvironment. Both studies demonstrated the functional delivery of respective therapeutics to intracranial GBM models, as validated by histological evidence of agent accumulation at tumor sites and significant improvements in survival, suggesting successful navigation of the BBB for therapeutic effect. A key strength of Zhu's study lies in its engineering of exosomes for targeted, non-cellular delivery, and its comparative testing against free PTX and non-functionalized exosomes. Bak's work is notable for its use of actively tumor-homing therapeutic cells that deliver gene therapy with spatial precision and extended activity. Together, these studies demonstrate two different approaches to stem cell-based therapy in GBM, each offering advantages that overcome current delivery challenges and highlighting key areas for further research. These stem cell-based delivery strategies are summarized in **Figure 1**. The key delivery platforms, therapeutic cargo, and mechanisms of action are summarized in **Table 1**.

Stem Cell-Based Drug and Gene Delivery

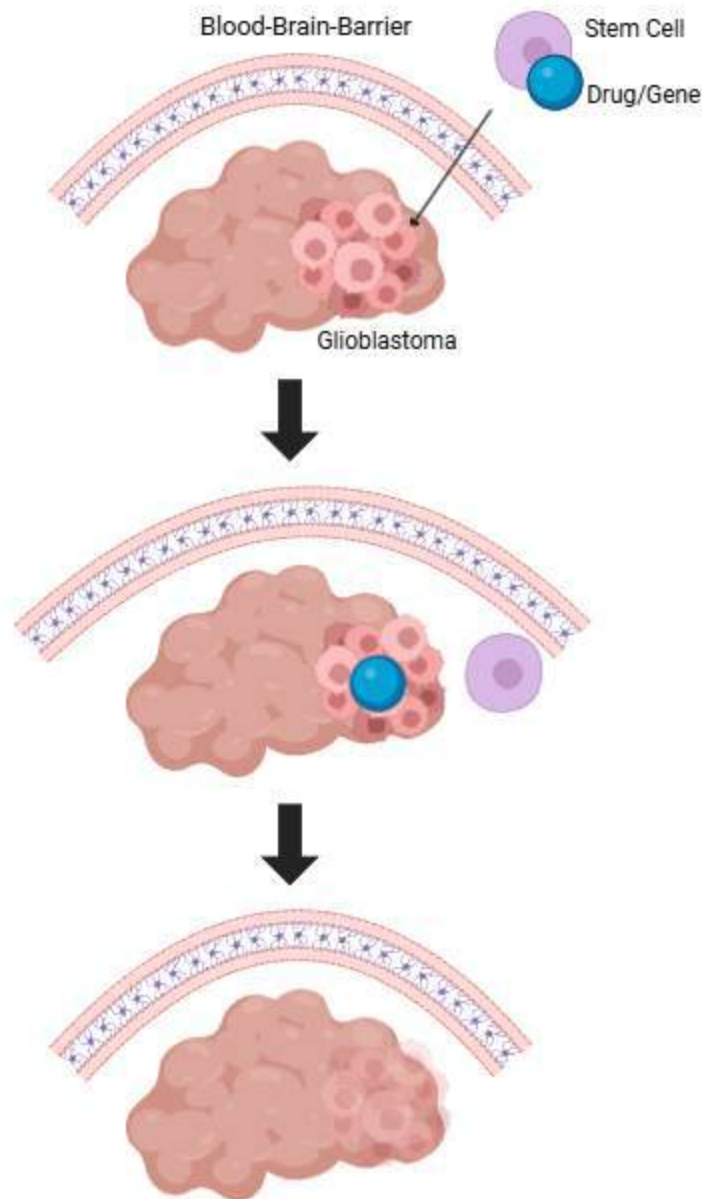


Figure 1: Stem Cell-based Drug and Gene Delivery: Stem cell-based drug and gene delivery in glioblastoma. Top) Engineered stem cells carrying therapeutic drugs or genes cross the blood–brain barrier and migrate toward glioblastoma tissue. Bottom) Release of the therapeutic cargo within the tumor induces tumor cell death.

Stem Cell-Based Immunomodulation in GBM

Beyond their regenerative and delivery capacities, stem cells are evolving for their potential to modulate the immune system in cancer therapy. In glioblastoma (GBM), immunotherapies face significant challenges due to the tumor’s suppressive microenvironment and the protective blood-brain barrier (BBB). Applications of MSCs to leverage the immune responses within this complex environment have attained prior attention (Nowak, 2021).

Strategies focused on activating immune cells, restoring antigen presentation, and counteracting the tumor's immune evasion mechanisms using MSCs have emerged as a potential avenue to address GBM's resistance to immune-mediated treatments.

Vogiatzi et al addressed a critical barrier in glioblastoma (GBM) treatment: the immunosuppressive tumor microenvironment, which renders immunotherapies largely ineffective (Vogiatzi, 2023). While immune checkpoint inhibitors and cytokine-based strategies have shown efficacy in other cancers, GBM's local immune evasion and the blood-brain barrier (BBB) limit their success. To overcome this, the study engineered allogeneic murine MSCs to serve as immunomodulatory delivery vehicles that can survive and directly alter the tumor microenvironment. The MSCs lacking interferon- α/β receptor 1 (IFNAR1), which makes them resistant to interferon- β (IFN β)-induced cell death, could be further engineered to secrete IFN β and a single-chain variable fragment that blocks PD-1 (scFv-PD1). This dual payload was designed to activate cytotoxic T cells while inducing immunogenic changes in GBM cells. This approach increased their survival to immune attack through enhanced PD-L1 expression and IFN β -mediated signaling. *In vitro*, IFN β secretion from engineered MSCs upregulated PD-L1 expression on GBM cells, directly reducing the viability of tumor cells (Vogiatzi 2023). Co-cultures with murine T cells confirmed that scFv-PD1 enhanced T cell activation and tumor regression, with combination therapy showing synergistic effects. *In vivo*, immunocompetent C57BL/6 mice with orthotopic GBM xenografts underwent intracranial implantation of scaffold-encapsulated engineered MSCs following surgical tumor resection. Mice treated with dual-expressing MSC KO-IFN β -scFv-PD1 cells exhibited significantly reduced tumor growth and prolonged survival, achieving a 40% survival rate at 60 days, compared to minimal survival in control groups, which included mice receiving unmodified MSCs, MSCs expressing only scFv-PD1, or no treatment. These findings provide evidence that the engineered MSCs not only reached the tumor site post-resection but also remained active and functionally altered the immune landscape, delivering a precise immunotherapy across the BBB. These findings present a potentially effective strategy for leveraging stem cells to modulate immune activity in GBM.

Mohme et al demonstrated that locally delivered interleukin-expressing MSCs reprogram the immunosuppressive GBM microenvironment to support a durable anti-tumor immune response (Mohme, 2020). While MSCs are known for their tumor-homing ability, their application in GBM immunotherapy remains limited by the challenge of generating sustained, targeted immune activation within the brain (Nowak, 2021). The study hypothesized that genetically engineering MSCs to coexpress interleukin-12 (IL-12), a potent activator of T and NK cells, and interleukin-7 (IL-7), a cytokine critical for T-cell survival and expansion, would promote a cytotoxic immune response against GBM. Modified MSCs expressing both the cytokines stably (MSC^{IL7/12}) and administered directly into orthotopic, immunocompetent C57BL/6 mouse models bearing GL261 or CT2A gliomas were utilized to test this hypothesis. Tumor progression was assessed using MRI, and immune responses were evaluated through flow cytometry. Local administration of MSC^{IL7/12} significantly suppressed tumor growth and led to complete tumor regression in up to 50% of treated mice, with survivors exhibiting long-term

protection upon tumor rechallenge. This study provides evidence that MSC-mediated cytokine delivery remodels the GBM immune landscape, resulting in both tumor clearance and immunological memory. By demonstrating localized immune activation and durable systemic protection, the study by Mohme et al provided a compelling foundation for advancing stem cell-based immunotherapy into clinical trials.

Overall, Mohme et al engineered mesenchymal stem cells (MSCs) to secrete interleukin-12 (IL-12), a cytokine that activates cytotoxic T cells and natural killer (NK) cells, aiming to boost anti-tumor immunity within glioblastoma. Their approach led to increased infiltration of CD8⁺ T cells and enhanced tumor-specific immune responses in both murine models and patient-derived xenografts. In contrast, Vogiatzis et al utilized MSCs to deliver oncolytic viruses that selectively infect tumor cells, resulting in tumor lysis, simultaneously releasing danger signals that attract and activate immune cells. This strategy induced a more localized immune activation; however, it failed to establish systemic immunity explicitly. Work induces a broader reprogramming of immune cell composition, evident by a CD8-dominated tumor response and increased T-cell receptor diversity. Importantly, both studies validate that MSCs serve as immunologically active vehicles capable of reshaping GBM's immune landscape. Notably, Mohme et al established durable systemic immunity, suggesting that cytokine-mediated strategies may be more effective for lasting tumor rejection. Together, these studies demonstrate complementary approaches to MSC-based immunotherapy, focusing on overcoming immune inhibition and sustaining anti-tumor immunity, both of which are critical in addressing GBM resistance to current immune interventions. The immunomodulatory mechanisms of engineered stem cells are illustrated in **Figure 2**.

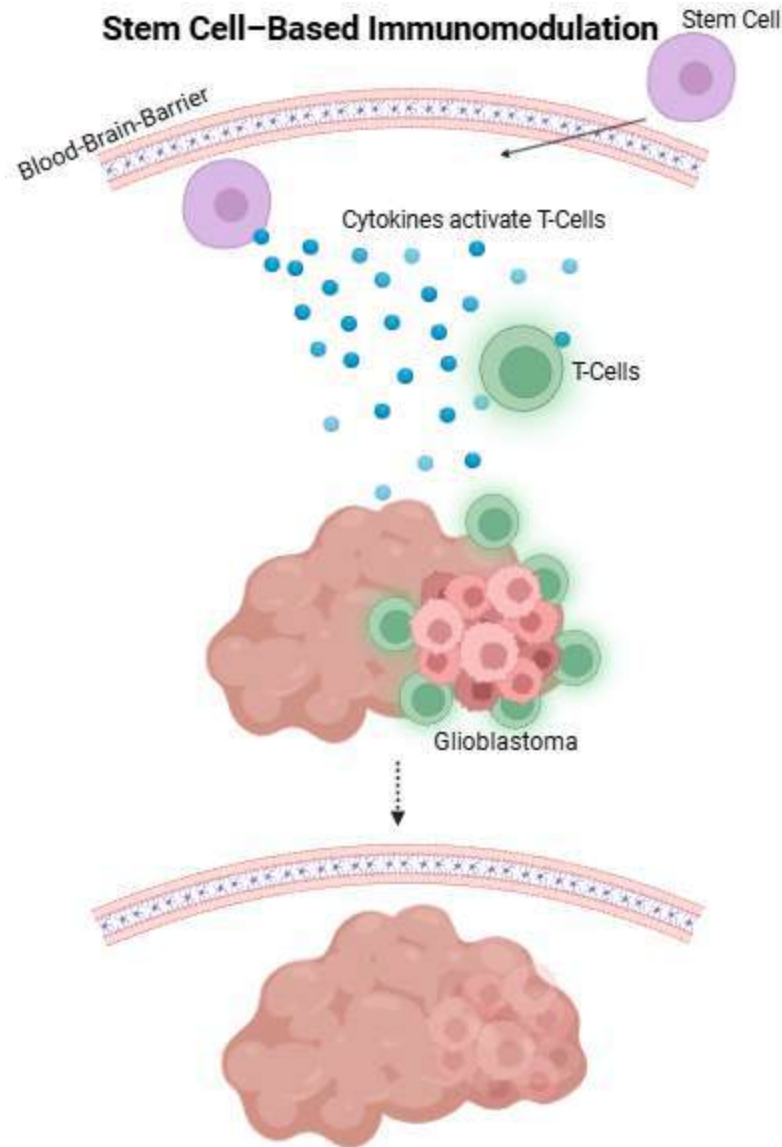


Figure 2: Stem Cell-Based Immunomodulation. Immunomodulatory effects of engineered stem cells in glioblastoma. Top) Stem cells secrete cytokines that activate cytotoxic T cells. Bottom) Activated immune cells infiltrate the tumor and promote glioblastoma cell death.

A comparison of immunomodulatory stem cell strategies and their immune targets is provided in **Table 2**.

Targeting Glioblastoma Stem Cells

While conventional glioblastoma (GBM) therapies often target the bulk of the tumor, they frequently fail to eliminate a small but critical subpopulation of tumor cells with stem-like properties, known as cancer stem cells (CSCs) (Biserova, 2021). These glioblastoma stem cells (GSCs) have self-renewal abilities, tumor-initiating capacity, and innate resistance to chemotherapy and radiation, making them key drivers of recurrence with poor prognosis (Gimple, 2019). Unlike differentiated tumor cells, GSCs are sustained by specialized transcriptional and epigenetic programs that preserve their stem-like properties and enable their survival under therapeutic stress. As a result, targeting GSCs has become a key focus in GBM

research, with efforts directed at disrupting their survival mechanisms and impairing their ability to self-renew. Across studies, strategies to eliminate GSCs range from inducing selective cytotoxicity to interfering with the molecular machinery that maintains their identity. These approaches signify a shift toward therapies that not only decrease tumor burden but also target the cellular root of GBM's treatment resistance.

Magrath et al addressed GSCs' resistance to conventional treatments as a driving factor for recurrence. The study investigated whether neural stem cells (NSCs) could be engineered to deliver a targeted cytotoxic signal directly to GSCs. NSCs were chosen owing to their inherent migration capacity toward tumor sites within the brain, making them promising vehicles for localized therapy (Li, 2023). The study modified human NSCs to secrete TRAIL (tumor necrosis factor-related apoptosis-inducing ligand), a protein that activates death receptors (DR4 and DR5) and induces apoptosis preferentially in tumor cells (Gampa 2023). The findings confirmed that patient-derived GSC cultures expressed high levels of DR4/DR5 and were sensitive to TRAIL treatment upon exposing the cells to recombinant TRAIL in vitro and measuring reduced cell viability. Further, the study employed a GSC-driven orthotopic xenograft model, in which patient-derived GSCs were implanted into the brains of immunocompromised mice, to closely mimic human tumor biology and recurrence patterns. TRAIL-secreting NSCs were implanted near the tumor site. Intravital imaging revealed that the NSCs migrated toward the tumor mass and released TRAIL locally. The treatment reduced tumor size, delayed disease progression, and significantly extended survival compared to controls. While apoptosis was inferred from the known mechanism of TRAIL and the sensitivity of GSCs in vitro, the in vivo findings were based on tumor regression and survival outcomes. Together, these results demonstrated that NSC-mediated TRAIL delivery effectively suppressed GSC-driven tumor growth in a clinically relevant model, highlighting its potential as a vehicle-based therapeutic strategy to overcome GSC-associated resistance in GBM.

Mitchell et al investigated the role of WDR5, a chromatin-associated protein (Aho, 2019), in sustaining the stem-like identity of glioblastoma cancer stem cells (CSCs) (Mitchell, 2023). CSCs depend on tightly controlled epigenetic programs to sustain self-renewal and resist differentiation, yet the specific chromatin regulators responsible for this state in glioblastoma remain understudied. WDR5 was selected because it is a core component of the WRAD complex, which facilitates histone H3 lysine 4 trimethylation (H3K4me3) at promoters of genes essential for stem cell maintenance. Also, WDR5 is critical for normal and malignant stem cell biology. Recognizing that CSCs are essential drivers of glioblastoma progression and recurrence due to their epigenetic plasticity and ability to self-renew (Biserova, 2021), WDR5 supports transcriptional programs necessary for maintaining these stem-like traits, including the capacity to self-renew, resist differentiation, and form new tumors as demonstrated using the small-molecule WDR5 inhibitor, C16, across multiple patient-derived glioblastoma organoid lines. These organoids preserve tumor heterogeneity and stem cell compartments, as confirmed by histological analysis and immunostaining for stemness markers, including Sox2. Treatment with C16 reduced organoid viability and selectively depleted the Sox2⁺ CSC population without

broadly affecting non-stem tumor cells as determined by flow cytometry. Further mechanistic studies revealed that WDR5 inhibition disrupted the WRAD complex, leading to the loss of H3K4me3 marks at the promoters of stemness-related genes, including SOX2 and OLIG2. This loss of activating methylation was associated with reduced transcription of stemness genes and a corresponding shift toward a more differentiated, less tumorigenic state. In vivo, WDR5 inhibition reduced the ability of tumors to initiate and increased survival in mice with xenografts. However, single-cell RNA sequencing revealed that CSCs exhibited adaptive transcriptional responses to WDR5 loss, including upregulation of alternate epigenetic regulators. These important findings suggest that while WDR5 inhibition disrupts a key dependency, CSCs activate compensatory pathways that undermine long-term therapeutic efficacy. Overall, this study identifies WDR5 as a central epigenetic dependency in glioblastoma CSCs and highlights the promise of targeting chromatin regulators to disrupt tumor-initiating potential.

The studies by Magrath et al and Mitchell et al target glioblastoma cancer stem cells (CSCs), but they pursue fundamentally different therapeutic strategies. One focuses on directly killing CSCs, while the other targets disrupting the epigenetic mechanisms that sustain their identity. Magrath et al engineered NSCs to secrete TRAIL, a ligand that engages DR4 and DR5 death receptors on CSCs to trigger apoptosis (Gampa, 2023). Their approach capitalized on NSCs' tumor-homing properties to achieve localized delivery, which not only reduced tumor burden but also limited off-target effects. Mitchell et al used a small-molecule inhibitor, C16, to block WDR5, a chromatin-modifying protein essential for maintaining CSC stemness programs. This strategy impaired the transcriptional machinery that sustains CSC identity, resulting in functional suppression even in cells that might otherwise evade direct cytotoxicity. Both studies used patient-derived models: orthotopic xenografts and organoids, which preserved tumor heterogeneity and CSC subpopulations. This is a strength, as it allows the therapies to be evaluated in systems that more closely mimic the diversity of cell types, microenvironmental influences, and treatment resistance mechanisms seen in human GBM, increasing the translational promise. However, the therapies differed in durability. TRAIL-based strategy demonstrated sustained CSC elimination with limited resistance, whereas Mitchell et al demonstrated adaptive changes in transcription factor networks that could facilitate therapeutic escape. Together, these studies reveal two complementary strategies to overcome treatment-resistant cells: one that induces CSC death through apoptosis, and another that weakens CSC function by reprogramming their epigenetic state. The two major strategies for targeting glioblastoma cancer stem cells are shown in **Figure 3**. Molecular goals and outcomes of CSC-directed therapies are summarized in **Table 3**.

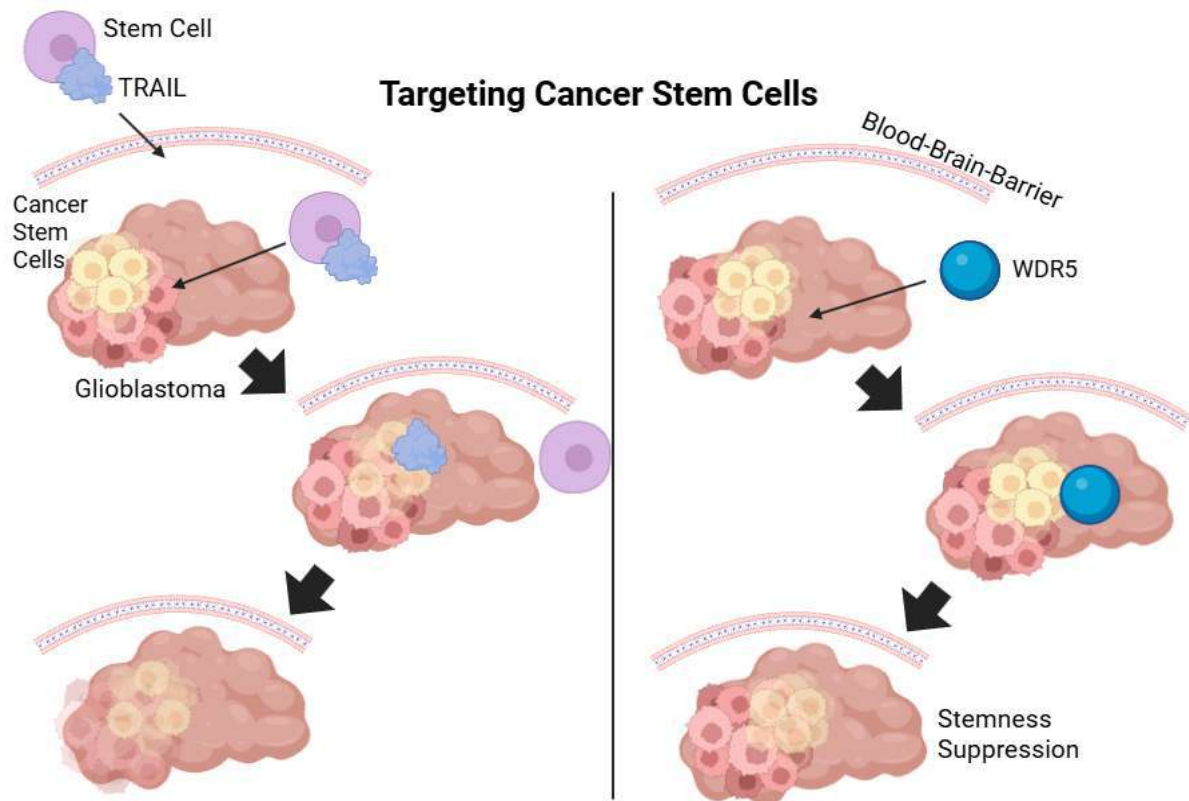


Figure 3: Targeting Cancer Stem Cells. Stem cell–based targeting of glioblastoma cancer stem cells (CSCs). Left) Stem cells engineered to release TRAIL induce apoptosis in CSCs. Right) Targeting WDR5 suppresses CSC stemness, reducing tumor survival and regrowth.

Table 1: Stem Cell-Based Drug and Gene Delivery

Stem Cell Type	Therapeutic Cargo	Mechanism of Action	Advantages
ESC-derived exosomes	Paclitaxel (PTX)	Cross BBB, tumor-targeted release	Enhanced tumor-specific uptake
hESC-derived MSC-like cells	HSV-TK + GCV	Prodrug activation leads to DNA damage	Scalable gene therapy
Neural stem cells	TRAIL	Induces apoptosis in GBM cells	Selective tumor killing cells

Table 2: Stem Cell–Mediated Immunomodulation in Glioblastoma

Engineered Stem Cell	Secreted Factor	Immune Target	Therapeutic Effect
MSC-IL7/IL12	IL-7, IL-12	T cells, NK cells	Sustained anti-tumor immunity
MSC-IFN β + scFv-PD1	IFN β , PD-1 blocker	T cells	Reverses immune

			suppression
Cytokine-secreting MSCs	Multiple cytokines	Tumor microenvironment	Increased immune infiltration

Table 3: Targeting Glioblastoma Cancer Stem Cells (CSCs)

Target Strategy	Molecular Target	Effect on CSCs	Outcome
TRAIL delivery	DR4 / DR5 receptors	Induces apoptosis	Reduced CSC survival
WDR5 inhibition	Epigenetic regulator	Stemness Suppression	Decreased tumor regrowth

Discussion

Stem cell-based therapies for glioblastoma aim to overcome the BBB and improve targeted delivery, immune activation, and cancer stem cell removal. A central unifying theme across these strategies is the ability of stem cells to cross or bypass the blood-brain barrier (BBB), a major obstacle that prevents most conventional drugs from effectively reaching tumor sites (Alahmari 2021; Wu 2021). Stem cells' tumor-homing capacity allows them to migrate through brain tissue and localize to glioblastoma, functioning as vehicles that can deliver therapeutic payloads to areas unreachable by systemic chemotherapies (Nowak 2021). Beyond delivery, immunomodulatory approaches are designed to enhance immune activation by modifying the tumor microenvironment to aid anti-tumor immunity, countering GBM immunosuppressive environments (Mohme 2020; Vogiatzi 2023). Finally, targeting GSCs represents an attempt to directly eliminate the self-renewing cell populations responsible for recurrence and resistance (Gimple 2019; Biserova 2021). Together, these three strategies capture the promising therapeutic process of stem cell-based approaches in glioblastoma.

Delivery approaches using mesenchymal and embryonic stem cell-derived systems excel at transporting drugs or genes into tumors. Additionally, immunomodulatory approaches focus on altering the tumor microenvironment to improve anti-tumor responses. Cancer stem cell-targeting strategies differ in that they directly eliminate the root population that drives recurrence. A major similarity across all strategies is the reliance on stem cells' tumor-homing ability, yet their methods differ in mechanism and clinical readiness (Omuro 2013; Aldape 2015). Delivery systems have shown encouraging preclinical results; however, they face challenges of safety and consistency in human patients. For example, a Phase I clinical trial using genetically modified neural stem cells to deliver a prodrug-converting enzyme to patients with glioblastoma (NCT02015819) demonstrated feasibility. Still, the observations highlighted safety concerns, such as potential for immune-related toxicities and variable persistence of stem cells across patients. Specifically, MSC- and ESC-derived exosomes have been shown to successfully carry chemotherapy agents such as paclitaxel into glioblastoma in vivo mouse models (Bălașa 2020; Zhu 2019), while engineered neural stem cells have been used to deliver prodrug-converting

enzymes (Bak 2011). These findings validate the potential of stem cell platforms for therapeutic delivery, but safety concerns remain, particularly the possibility that stem cells could promote tumor growth by differentiating into tumor-associated stromal elements or secreting factors that support glioblastoma proliferation (Nowak 2021).

Immunomodulation offers the potential of stable responses, though it risks off-target immune effects. For example, stem cells engineered to secrete interleukins such as IL-12 or IFN- β recruit cytotoxic T cells and reduce tumor volume in animal models (Mohme, 2020; Vogiatzi, 2023). However, they could provoke systemic inflammation or autoimmune reactions under unregulated circumstances (Nowak, 2021). CSC targeting is conceptually powerful; however, it remains in early stages. Preclinical studies with TRAIL-expressing NSCs demonstrate the selective killing of GSCs (Gampa 2023). Inhibition of epigenetic regulators, such as WDR5, has shown promise in reducing GSC self-renewal (Mitchell 2023; Aho 2019). However, these strategies are limited by challenges in ensuring specific delivery to CSCs without affecting normal neural stem cells, and uncertainty about long-term safety in vivo. The field of stem cell-based therapies for glioblastoma has explored both neural stem cell (NSC) and dendritic cell (DC)-based approaches. Genetically modified NSCs expressing cytosine deaminase, combined with the prodrug flucytosine and leucovorin, have been tested in patients with recurrent high-grade gliomas (NCT02015819), demonstrating safety and tolerability in early-phase trials. Immunotherapeutic strategies using DCs loaded with autologous or allogeneic glioma stem cells have also been investigated (NCT00846456; NCT01171469), showing favorable safety records and evidence of obtaining tumor-specific immune responses. While these studies confirm that stem cell-based interventions are feasible in humans, challenges remain, such as limited effectiveness in controlling tumor growth, variability in patient responses, and the complexity of manufacturing these therapies. Further research is required to improve their therapeutic effectiveness and standardize treatment protocols.

Emerging advancements in drug delivery and immunomodulatory approaches by engineering stem cells with greater specificity and combining them with established therapies offer synergistic effects. Current efforts include expressing targeting ligands, such as c(RGDyK) peptides, to exosome membranes to enhance binding to glioblastoma cells (Zhu 2019), as well as engineering stem cells with inducible promoters to control therapeutic release exclusively at the tumor site (Li 2023). Integrating these platforms with existing modalities: for example, pairing stem cell-delivered oncolytic viruses with radiation to increase viral replication, or combining stem cell immunomodulation with immune checkpoint inhibitors to amplify T cell activity, is promising (Vogiatzi 2023; Wirsching 2016). This combinatorial strategy aims to leverage the strengths of both novel and standard therapies, while overcoming their limitations when used alone.

Nevertheless, the field faces obstacles, including tumor heterogeneity, incomplete understanding of stem cell-tumor interactions, and ethical concerns in the use of embryonic stem cells. Tumor heterogeneity means that glioblastoma within a single patient can carry diverse genetic subclones, some of which may evade stem cell-delivered treatments, thereby dampening

long-term efficacy (Aldape 2015; Wu 2021). Finally, embryonic stem cell-derived therapies, while powerful, raise ethical concerns owing to the destruction of embryos and face regulatory hurdles that limit their scalability (Biswas 2007). These ongoing issues underscore the need for careful preclinical evaluation and rigorous ethical frameworks before the broad implementation of stem cell-based therapies for glioblastoma.

Future research warrants focus on enhancing the precision of delivery and ensuring safety through rigorous preclinical and clinical testing. Although obstacles remain, the potential of stem cell-based strategies lies in their unique ability to reach invasive tumor regions and address recurrence, making them a promising treatment option for GBM. Overall, this review found that stem cell-based therapies offer potential innovative solutions to GBM treatment challenges by enhancing delivery, modifying the tumor microenvironment, and eliminating resistant cancer stem cell populations. The field currently shows potential, with promising preclinical evidence but limited clinical application (Omuro 2013; Wirsching 2016). For instance, delivery approaches have repeatedly shown that stem cells can carry drugs across the BBB (Zhu 2019; Bak 2011), immunomodulatory methods have reshaped tumor immunity in animal models (Mohme 2020; Vogiatzi 2023), and GSC targeting has reduced recurrence in xenograft systems (Magrath 2020; Mitchell 2023). However, translation to humans remains limited, with only a few trials and early safety data available (Wu 2021). Future directions require improving safety through tighter regulation of stem cell behavior, enhancing targeting specificity with ligands and gene circuits (Zhu 2019; Li 2023), and integrating stem cell therapies with standard care to maximize efficacy (Aldape 2015). With continued refinement, stem cell-based approaches pose the potential to transform clinical outcomes for GBM patients.

Conclusion

Stem cell-based therapies provide a potential strategy to overcome key challenges in glioblastoma treatment, including limited drug delivery across the blood–brain barrier, immune suppression, and tumor recurrence. By leveraging tumor-homing properties, stem cells can act as delivery vehicles, immunomodulators, and targeted approaches against therapy-resistant cancer stem cells. While clinical translation remains limited, strong preclinical evidence suggests these approaches could significantly improve glioblastoma outcomes with further refinement.

Works Cited

- Aho, Emily R., et al. "Targeting WDR5: A Winning Anti-Cancer Strategy?" *Epigenetics Insights*, vol. 12, 2019, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/31360909/.
- Alahmari, A. "Blood–Brain Barrier Overview: Structural and Functional Correlation." *Neural Plasticity*, 2021, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/34912450/.
- Aldape, Kenneth, et al. "Glioblastoma: Pathology, Molecular Mechanisms and Markers." *Acta Neuropathologica*, vol. 129, no. 6, 2015, pp. 829–848. U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/25943888/.
- Bak, Xiao Ying, et al. "Human Embryonic Stem Cell-Derived Mesenchymal Stem Cells as Cellular Delivery Vehicles for Prodrug Gene Therapy of Glioblastoma." *Human Gene Therapy*, vol. 22, no. 11, 2011, pp. 1365–1377. U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/21425958/.
- Bălașa, Andrei, et al. "The Involvement of Exosomes in Glioblastoma Development, Diagnosis, Prognosis, and Treatment." *Brain Sciences*, vol. 10, no. 8, 2020, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/32823792/.
- Biswas, A., and R. Hutchins. "Embryonic Stem Cells." *Stem Cells and Development*, vol. 16, no. 3, 2007, pp. 425–437. U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/17521233/.
- Biserova, K., et al. "Cancer Stem Cells: Significance in Origin, Pathogenesis and Treatment of Glioblastoma." *Cells*, vol. 10, no. 4, 2021, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/33799798/.
- Gampa, S. C., et al. "Nano-TRAIL: A Promising Path to Cancer Therapy." *Cancer Drug Resistance*, vol. 6, 2023, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/37065863/.
- Gimple, Ryan C., et al. "Glioblastoma Stem Cells: Lessons from the Tumor Hierarchy in a Lethal Cancer." *Genes & Development*, vol. 33, no. 11–12, 2019, pp. 591–609. U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/31160393/.
- Li, G., et al. "Targeting the Neural Stem Cells in Subventricular Zone for the Treatment of Glioblastoma: An Update from Preclinical Evidence to Clinical Interventions." *Stem Cell Research & Therapy*, vol. 14, 2023, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/37170286/.
- Magrath, James W., et al. "In Vitro Demonstration of Salinomycin as a Novel Chemotherapeutic Agent for the Treatment of Sox2-Positive Glioblastoma Cancer Stem Cells." *Oncology Reports*, vol. 44, no. 1, 2020, pp. 57–66. U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/32627023/.
- Mitchell, Kevin, et al. "WDR5 Represents a Therapeutically Exploitable Target for Cancer Stem Cells in Glioblastoma." *Genes & Development*, vol. 37, no. 5–6, 2023, pp. 328–345. U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/36732025/.

- Mohme, Martin, et al. “Local Intracerebral Immunomodulation Using Interleukin-Expressing Mesenchymal Stem Cells in Glioblastoma.” *Clinical Cancer Research*, vol. 26, no. 11, 2020, pp. 2626–2639. U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/31988196/.
- Nowak, Barbara, et al. “Mesenchymal Stem Cells in Glioblastoma Therapy and Progression: How One Cell Does It All.” *Biochimica et Biophysica Acta (BBA) – Reviews on Cancer*, vol. 1876, no. 2, 2021, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/34144129/.
- Omuro, Antonio, and Lisa DeAngelis. “Glioblastoma and Other Malignant Gliomas: A Clinical Review.” *JAMA*, vol. 310, no. 17, 2013, pp. 1842–1850. U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/24193082/.
- Vik-Mo, Einar Osland, et al. “Therapeutic Vaccination against Autologous Cancer Stem Cells with mRNA-Transfected Dendritic Cells in Patients with Glioblastoma.” *Cancer Immunology, Immunotherapy*, vol. 62, no. 9, 2013, pp. 1499–1509, <https://doi.org/10.1007/s00262-013-1453-3>.
- Vogiatzi, Ioanna, et al. “Allogeneic Stem Cells Engineered to Release Interferon β and scFv-PD1 Target Glioblastoma and Alter the Tumor Microenvironment.” *Cytotherapy*, 2023, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/38852095/.
- Wirsching, Hans-Georg, et al. “Glioblastoma.” *Handbook of Clinical Neurology*, vol. 134, 2016, pp. 381–397. U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/26948367/.
- Wu, Wen, et al. “Glioblastoma Multiforme (GBM): An Overview of Current Therapies and Mechanisms of Resistance.” *Pharmacological Research*, vol. 171, 2021, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/34302977/.
- Zhu, Qingwei, et al. “Embryonic Stem Cells-Derived Exosomes Endowed with Targeting Properties as Chemotherapeutics Delivery Vehicles for Glioblastoma Therapy.” *Advanced Science*, vol. 6, no. 17, 2019, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/30937268/.
- “Genetically Modified Neural Stem Cells, Flucytosine, and Leucovorin for Treating Patients with Recurrent High-Grade Gliomas.” *ClinicalTrials.gov*, 2025, www.clinicaltrials.gov/study/NCT02015819.
- “Vaccination with Dendritic Cells Loaded with Brain Tumor Stem Cells for Progressive Malignant Brain Tumor.” *ClinicalTrials.gov*, 2025, www.clinicaltrials.gov/study/NCT01171469.

The Effect of Minimum Wage Increases on the Local Economy: A Case Study of Seattle

By Isaac Sedighi, Rocco Graziosi

I. Introduction

Minimum wage is the lowest possible wage that an employer can pay an employee. Since the concept of the minimum wage was first established in 1938, there has been a ceaseless debate over whether it is at an optimal level. Politicians and economists have mainly considered its impact on the economy, which includes, but is not limited to, unemployment, inflation, and small businesses. The unemployment rate is the proportion of the labor force that is not employed. The metric reflects only the people who are actively seeking work, not the people who do not want to work. A decline in consumers' purchasing power, along with the general increase in the cost of living, is called inflation. It can be measured using the Consumer Price Index (CPI), which tracks the average change in prices of goods and services purchased by households over time. To classify a business as small, it must have fewer than 500 employees, and it is typically a for-profit organization. Small businesses account for approximately 99.9% of businesses in the United States, 43.5% of the national GDP, and two out of every three new jobs created (USAFacts, 2024).

In this paper, Seattle is used as the primary case study, with some data also covering other nearby cities in Washington state, such as Tacoma, Bellevue, and Kent. As of 2025, Seattle has the highest minimum wage of any major city in the United States, with expected increases in the coming years. But the city's yearly minimum wage hikes are not a recently started endeavor. Since 2015, Seattle has been increasing its minimum wage at a Compound Annual Growth Rate (CAGR) of approximately 6.5%. This is because of the minimum wage ordinance signed in 2014. It essentially said that minimum wage will rise to \$15/hr through a multi-year phase-in, then continue to adjust based on the cost of living. Large employers had a 2-3 year period, depending on employee benefits, until the Seattle minimum wage hit \$15 after the ordinance. On the other hand, the rise to \$15/hr was more spread out, time-wise, for small businesses because of their smaller profit margins and lack of spare capital. At the beginning of 2025, Seattle set the minimum wage to \$20.76/hr for *all* businesses. Future minimum wage increments based on inflation will continue to be universal to all employers, regardless of company size, tip credits, or employee benefit credits.

Seattle's Minimum Wage

Small Employers Benefits/Tips?		Large Employers Benefits?	
Yes	No	Yes	No
\$11.00	\$13.00	\$13.50	\$15.00
\$11.50	\$14.00	\$15.00	\$15.45
\$12.00	\$15.00	\$16.00	
\$13.50	\$15.75	\$16.39	
\$15.00	\$16.69	16.69	
\$15.75	\$17.27	\$17.27	
\$16.50	\$18.69	\$18.69	
\$17.25	\$19.97	\$19.97	
\$20.76		\$20.76	

Seattle Office of Labor Standards
Figure 1.1

The purpose of this study is to answer the following question: How have Seattle's minimum wage increases affected the city's local economy? The objective is to determine the impacts on consumers, workers, and small businesses by analyzing related qualitative observations and quantitative metrics. This paper argues that the rise of minimum wage in Seattle has significantly increased labor costs for small businesses, leading to price increases for consumers and producers, business closures, and reduced hiring, all of which will slow local economic growth and elevate unemployment over time.

II. Economic Background

In a free labor market, the demand for labor and labor supply naturally meet at an equilibrium wage value. However, a government-imposed minimum can create a disequilibrium in the job market.

As seen in the labor market supply and demand graph (Figure 2.1), the Theory of Labor Demand postulates that as minimum wage increases (Figure 2.1 - W_2), the quantity of labor demanded decreases, meaning fewer job opportunities. This is due to the fact that the increased labor costs disallow employers from hiring the original number of workers with the original amount of money. On the other hand, a decrease in minimum wage (Figure 2.1 - W_3) should theoretically increase the quantity of jobs because it costs less to hire the original number of workers. Furthermore, depending on the elasticity of labor demand (Equation 2.1), firms will

react differently to minimum wage changes. For example, for a firm with a high elasticity of labor demand, generally a small business, the quantity of labor demanded will decrease much more after a minimum wage increase compared to a firm with a low elasticity of labor demand, generally a large business. This negative correlation between the size of business and the elasticity of labor demand is because smaller businesses have less spare capital and a tighter budget. Hence, they cannot absorb the impact of a minimum wage increase as freely as a large corporation. Therefore, the greater the elasticity of labor demand, the greater the loss of opportunities for employment after a minimum wage increase.

The Theory of Labor Supply explains that as minimum wage increases (Figure 2.1 - W_2), the supply of available workers increases because they are more willing to work if the wage is higher. This is because the opportunity cost is worth more than the value of them not working, also called the Substitution Effect. However, the Income Effect must also be taken into account. The Income Effect suggests that some workers may reduce hours if they can maintain their same original income. Additionally, according to the Theory of Labor Supply, a decrease in minimum wage (Figure 2.1 - W_3) leads to a decrease in the quantity of labor supplied. Essentially, dealing with the opportunity cost of having leisure time becomes more attractive than the income from work, leading to a decrease in workers willing to work. Regarding elasticity of labor supply (Equation 2.2), workers with higher elasticities will react more strongly to minimum wage increases than those with lower elasticities. Elasticity of labor supply can be affected by unemployment benefits, alternative employment opportunities, personal views (Substitution Effect), etc.

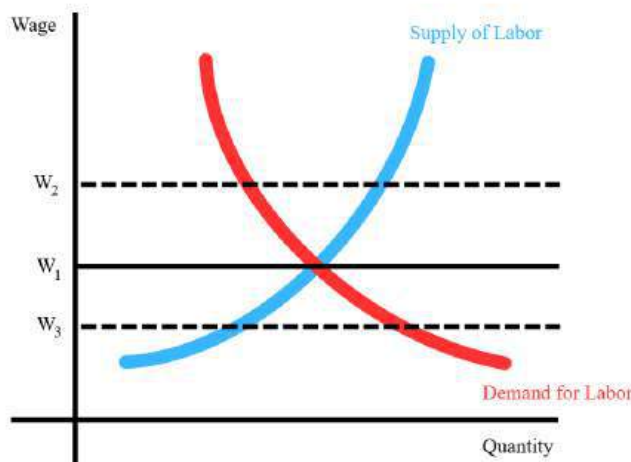


Figure 2.1: Labor Supply and Demand

$\frac{\% \text{ Change in Quantity of Labor Demanded}}{\% \text{ Change in Minimum Wage}}$

$\frac{\% \text{ Change in Quantity of Labor Demanded}}{\% \text{ Change in Minimum Wage}}$

Equation 2.1: Formula for Elasticity of Labor Demanded

$\frac{\% \text{ Change in Quantity of Labor Supply}}{\% \text{ Change in Minimum Wage}}$

$\frac{\% \text{ Change in Quantity of Labor Supply}}{\% \text{ Change in Minimum Wage}}$

Equation 2.2: Formula for Elasticity of Labor Supply

Together, the labor demand and the labor supply theories show the fundamental issue in the hands of politicians and economists concerning the minimum wage.

After a minimum wage increase, a small business cannot continue to function as it did prior to the change. Higher wages increase the cost of producing the goods and services that the business needs to produce. As labor costs rise, small businesses need to find a way to cover those costs. In order to maintain profit margins, small businesses often increase prices for consumers, which is called Cost-Push Inflation. One effect of the pass-through pricing is a phenomenon called a Wage-Price Spiral. As minimum wage increases, businesses increase the costs of goods/services to offset the increased labor costs. Therefore, the cost of living increases, and consumer purchasing power decreases. Then, the workers demand higher wages in order to continue to afford their original lifestyle, and the cycle continues.

III. Methodology

This paper will use Seattle as a case study to analyze the real-world impact of minimum wage increases on the local economy. Seattle was the first major U.S. city to pass a \$15 minimum wage law, prompting numerous universities and research institutions to study the effects of the increase, which resulted in numerous peer-reviewed documents and a wealth of collected data to explore. Therefore, there is significantly more data regarding the impacts of the minimum wage increases in Seattle compared to other cities with high minimum wages. Also, the city implemented the minimum wage increase through a phase-in, so observed impacts are less likely to occur due to economic shock, but more likely due to the policy itself. The phase-in was scheduled differently depending on the size of the business, allowing for an easier examination of how different-sized firms responded to the increase (Seattle Office of Labor Standards, 2025). Additionally, Seattle has a great deal of small businesses (2.55/100 small businesses per capita ratio), a strong labor market, and a diverse selection of sectors (tech, service, retail, etc.). All of this evidence supports the fact that Seattle has a comparable situation to other metropolitan areas if they were to increase minimum wage, making Seattle the ideal case study.

Both quantitative and qualitative data will be examined to provide a well-rounded study of Seattle. Quantitative data include, as mentioned in the Introduction, Consumer Price Index and Unemployment Rates, as well as business survival rates, profit margin trends, total compensation costs, etc. Qualitative data include surveys, interviews, news reports, case examples, etc., in order to identify adaptation strategies from and real-world effects on small businesses. Possible predicted adaptation strategies include modifying hours, hiring higher-skilled workers, automating processes, or increasing prices.

The sources used in this paper are official, academic, and local sources that ensure accurate and relevant information and conclusions. Official reports from the City of Seattle Office of Labor Standards, along with the U.S. Bureau of Labor Statistics, provide information on minimum wage policies, (un)employment trends, and more. Also, academic research from multiple universities and journals offers peer-reviewed articles on previous research done, as well as the economic theory pertaining to the study. Finally, local business and media reports supply qualitative evidence of impacts on small businesses, contextualizing the quantitative data.

Together, these sources ensure that the study's conclusions are defended by credible evidence and applicable to understanding Seattle's economic dynamics, and relevant in predicting the impacts on the local economy in other metropolitan areas after a minimum wage increase.

IV. Results and Analysis

In March of 2025, Seattle's total compensation was up approximately 5.8% year-over-year, which is large relative to the United States' overall of 3.4% (U.S. Bureau of Labor Statistics, 2025), and second to no other city in the country. Meaning that in Seattle, employers are paying their workers 5.8% more than they did last year, including both direct salary and indirect benefits. However, this number is moderated due to the fact that employee benefits are not as volatile as the actual salary being paid. The mean hourly wage in Seattle is \$43.16, whereas the national average is \$32.66 (U.S. Bureau of Labor Statistics, 2025), further revealing the undeniably high labor costs for firms.

The percentage of businesses that opened five years ago that are still open today is 41.1% in Seattle, and once again, the lowest and most extreme percentage nationally (Axios, 2025). Business closures in Seattle are at an all-time high because small businesses cannot afford the wages that the employees are required to receive. A particularly notable and widely known case study involving this situation is the closure of Bebo Waffle Shop in Seattle. Many news outlets reported that the restaurant could not sustainably handle the minimum wage increase and was forced to shut down (KOMO News, 2025; Eater Seattle, 2025; Fox13 Seattle, 2025). While large corporations are not highly affected, franchises are still facing the repercussions. Multiple McDonald's locations have closed in Seattle for this reason. Overall, the restaurant industry is the most strongly impacted due to the prevalence of low-skilled workers and low profit margins.

Small businesses are required to react after the minimum wage increases to try to keep their operations sustainable. The primary adaptation strategy for small businesses is reducing hires and limiting working hours for employees. For example, the manager at Cascadia Pizza in Seattle said he "had to let some people go" because of the increased labor costs (KNKX, 2025). He decided to cut multiple low-skilled full-time and part-time positions, primarily affecting the relatively new, inexperienced, and younger employees. "If you're going to be paying these premium wages, you want [your workers] to have premium skill sets," he said (KNKX, 2025).

Other common adjustments include modest price increases and operational innovations. At Geraldine's Counter Restaurant, they upped the menu prices by 10%, with another 10% likely soon following; they also added a 3% credit card surcharge to bills (KOMO News, 2025). Regarding operational innovations, many restaurants are trying to improve efficiency with their resources, whether it's using kiosks and robots in-store, adopting delivery models, streamlining menus, etc. Quyen Phan's Anchovies & Salt leases robots to bring food to tables, saving him "\$27,600 to \$30,000 per month" because he doesn't have to hire waiters (KOMO News, 2025). But this comes with obvious negative consequences; mainly that jobs are being lost, and often being lost to robots.

All these attempted adaptations may help keep businesses running, but they are reducing consumer affordability and employment opportunities, which likely limits long-term growth.

The high supply of workers will mean nothing due to firms' incredibly low demand for human labor because they are unable to afford it, and Seattle is already seeing the impacts of business closures and adaptations. In 2022, Seattle's unemployment rate was 3.2%; it has been on a slow and steady rise since then, and it is currently at 4.6% (U.S. Bureau of Labor Statistics, 2025). As wage growth continues to outpace productivity in businesses, there will be a continued rise in unemployment.

V. Discussion

Shortly, increased wages are leading to business closures, price increases, and the rise of unemployment in Seattle. The reduced hiring in restaurants is due to a highly elastic labor demand curve because of low profit margins. Also, the concept of cost-push inflation is clearly seen as Seattle restaurants are raising prices to combat higher payroll costs with increased revenues. Overall, the reaction of the local economy in Seattle aligns with major economic theory.

The reason for Seattle's initial high cost of living, which led to the minimum wage increases, was the rise of the technology field and the subsequent increase in housing prices. As companies like Microsoft and Amazon brought tens of thousands of high-wage technology workers into the city, the population boomed, and the 2010s decade ended with Seattle's population up 21% (Seattle Agent Magazine, 2021). Naturally, there would be an increased demand for housing, and there was. This was the primary contributor to the median house price in Seattle being approximately \$887,000 (Axios, 2025), which is almost double the median house price in America, approximately \$462,000 (Forbes, 2025). Prices in Seattle as a whole were already on the rise, and then the minimum wage ordinance was passed, making it increasingly difficult for businesses and individuals to keep up with the costs.

Despite the strong impact on Seattle's restaurant industry, plenty of Seattle's other businesses were relatively prepared for the wage increases. The city has a history of a progressive labor policy, which could imply that the reaction was less disruptive than it would be in other cities, perhaps being an impactful third party in this case study. Many employers had already been paying wages above the Seattle level, possibly either due to competitive pressure from the tech sector or just the labor policies of Seattle; regardless, the ordinance only formalized the pay levels that were already common in fields with a majority of high-skilled workers. Many firms had much greater flexibility when dealing with the minimum wage increases. As a result, the effect was concentrated primarily on the low profit margin, low wage industries- most notably the restaurant industry.

Our view is that the best minimum wage policy is different minimum wages for different business sizes. Evidently, large firms with greater profit margins are more resistant to minimum wage increases, whereas smaller businesses, with small profit margins, such as restaurants, often face the possibility of closure when minimum wage rises. Therefore, depending on factors such

as annual revenue, number of employees, profit margins, etc., larger businesses will have a greater minimum wage than small businesses. The Seattle case study largely supports the possible efficacy of this method, as the large technology businesses were essentially unaffected compared to the restaurants. Tailoring minimum wage would mitigate negative economic effects while still ensuring that workers receive fair compensation.

VI. Conclusion

The Seattle minimum wage ordinance, passed in 2014, which has significantly increased wages and overall labor costs, has strongly affected the local economy, particularly low-margin industries such as restaurants. Basic labor supply and demand theory is supported in the response from the Seattle economy. Small businesses have mostly reacted with pass-through costs, decreasing hours for workers, and limiting hiring in general, with some working on adopting operational innovations such as efficient staffing and automation. While these adaptations have kept some businesses open for the short term, long-term growth has been severely affected by rising unemployment and decreasing consumer affordability and spending.

The effect on individuals and small businesses has been predominantly negative. Individuals are having trouble affording goods such as housing, food, etc. Small businesses are being forced to adapt and cut costs in order to stay in business. However, the original plan of Seattle for a gradual phase-in with different minimum wages for different business sizes was effective in letting the businesses adjust, but ultimately only delayed the inevitable increase in closure rates for small businesses throughout the city.

Future research should investigate the effects of minimum wage increases in other cities and determine what attributes have the greatest influence on the impact of the increases. Other researchers should also compare their findings to ours to decide whether results are generalizable. Also, longer-term effects should be analyzed to conclude whether predictions in this study were correct.

Works Cited

- “Changing Compensation Costs in the Seattle Metropolitan Area — March 2025.” U.S. Bureau of Labor Statistics, Mar. 2025,
www.bls.gov/regions/west/news-release/2025/employmentcostindex_seattle_20250505.htm.
- Cheadle, Harry. “Why West Seattle’s Bebop Waffle Shop Went Viral in Conservative Media.” Eater Seattle, 21 Feb. 2025,
seattle.eater.com/2025/2/21/24369233/bebop-waffle-shop-minimum-wage-closure-viral-west-seattle-breakfast.
- Clarridge, Christine. “Why Seattle’s Housing Market Feels Frozen in Place.” Axios, 18 Nov. 2025,
www.axios.com/local/seattle/2025/11/18/seattle-housing-turnover-home-record-low-rising-listings-2025.
- Fitzpatrick, Anne. “Startup Survival Rates, Mapped.” Axios, 4 Nov. 2025,
www.axios.com/2025/11/04/new-business-survival-rates-map.
- Goddard, Amanda. “Why Was the Minimum Wage First Established?” History News Network, 13 Dec. 2016,
www.historynewsnetwork.org/article/why-was-the-minimum-wage-first-established.
- Hall, Alexandra. “Seattle Waffle Shop Owner Shuttters Business amid New \$20 Minimum Wage: ‘I’ve Cried Every Day.’” FOX Business, 3 Jan. 2025,
www.foxbusiness.com/media/seattle-waffle-shop-owner-shuttters-business-amid-new-20-minimum-wage-ive-cried-every-day.
- Moreno, Jonathan. “West Seattle Restaurant Owner Closes Doors as Rising Minimum Wage Strains Finances.” KOMO News, Jan. 2025.
- “Minimum Wage.” Seattle Office of Labor Standards, 2025,
www.seattle.gov/laborstandards/ordinances/minimum-wage.
- “Occupational Employment and Wages in Seattle-Tacoma-Bellevue.” U.S. Bureau of Labor Statistics, May 2024,
www.bls.gov/regions/west/news-release/OccupationalEmploymentAndWages_Seattle.htm.
- Reinhard, Taylor. “How Seattle’s Record-High Minimum Wage Has — and Hasn’t — Paid Off.” Interview by Nick Sanford, 31 Mar. 2025.
- “Seattle, WA Unemployment Rate.” YCharts, Aug. 2025,
ycharts.com/indicators/seattle_wa_unemployment_rate.
- “Seattle’s Minimum Wage.” Seattle Office of Labor Standards, 2025,
www.seattle.gov/documents/Departments/LaborStandards/2025_OLS_MW_MultiyearChart_FINAL.pdf.
- Tolentino, Stephanie. “Surging Egg Prices Prompt New Shopping Habits, Menu Changes in Seattle Area.” Interview by David Andre, 18 Feb. 2025.

- Treece, Aly J. Kiah. "Median Home Price by State: How Much Do Houses Cost?" Forbes, 18 Sept. 2025,
www.forbes.com/advisor/mortgages/real-estate/median-home-prices-by-state/.
- USAFacts Team. "What Role Do Small Businesses Play in the Economy?" USAFacts, 22 Nov. 2024, usafacts.org/articles/what-role-do-small-businesses-play-in-the-economy/.
- Wells, Logan. "Seattle Ended Decade with Faster Growth than Suburbs, Reversing 100-Year Trend." Seattle Agent Magazine, 9 Dec. 2021,
seattleagentmagazine.com/2021/12/09/seattle-ended-decade-with-faster-growth-than-suburbs-reversing-100-year-trend/.

Advancements in Machine Learning Applications for Earthquake Prediction By Alex Chen

Abstract

Predicting earthquakes remains one of the most persistent challenges in the field of geoscience. Traditional seismic projection models based on fault physics and statistical relationships from historical earthquake data fall short in reflecting the complexity in seismological patterns, failing to accurately forecast future seismological activity. In recent years, machine learning (ML) techniques have emerged as powerful tools to address these limitations, inherently being able to adapt and learn from growing volumes of data that conventional models fail to properly account for. This review paper examines key ML techniques and their possible applications. By assessing their current strengths, constraints, and practicality, this paper provides a comprehensive overview of the current state of the field.

Introduction

On January 17, 1994, the Los Angeles region was struck by a magnitude 6.7 earthquake lasting approximately 15 seconds, causing 57 fatalities, injuring thousands, and inflicting an estimated \$20 billion in damages. (U.S. Geological Survey [USGS], 1994). Despite being considered a moderate event on the Richter scale, the Northridge earthquake would come to be recognized as one of the costliest seismic events in U.S. history, underscoring the need for improved methods of predicting earthquakes. Having the technology for accurate predictions would enable appropriate preparation against occurrences, limiting the human and economic losses. However, traditional seismological techniques often fall short in forecasting earthquake occurrences and magnitudes (Michael & Langbein, 1993). Recently, the application of ML in identifying patterns and indicators of seismic activity has emerged as a promising approach to advancing seismic technology.

The applications of ML in earthquake prediction can be categorized into two different approaches to predicting earthquakes: forecasting and nowcasting (Field et al., 2015; Fox et al., 2021). Earthquake forecasts estimate a general probability of earthquakes occurring in an area over long periods of time (months to years) through analyzing data from past earthquakes and correlating attributes such as their magnitude, location, frequency, as well as fault-system relationships and seismic waveform characteristics (Field et al., 2015). On the other hand, nowcasting involves calculating the immediate probability of an earthquake occurring, providing an estimate of the current hazard level using data streams of ongoing seismicity and geodetic recordings (Fox et al., 2021; Feng & Fox, 2021). Both approaches utilize the ability of ML to assess large volumes of multiple types of data and identify subtle correlations between data points, while also improving and adapting their models as additional observations become available, often automating tasks that traditional statistical models are incapable of performing (Yavas et al., 2024).

Despite the promising results, the usage of ML in seismic prediction still faces several challenges. The quality of historical earthquake data and their availability remain issues, as

comprehensive datasets for larger earthquakes are limited, due to their infrequency. The reliability of data-driven models is also a concern, as there is a lack of transparency in their decision-making processes.

This paper aims to review research aimed at advancements in ML technology in the forecasting and nowcasting of earthquakes. The study surveys various ML techniques and how they've been employed, highlighting the current state of the field, and evaluating the future direction of ML-based seismic prediction technology.

Background

Over centuries, tectonic plates rub against each other, storing elastic energy along fault lines. Earthquakes occur when accumulated stress in the Earth's crust exceeds the strength of the surrounding rock, releasing energy that causes widespread destruction. California and much of the U.S. West Coast experience frequent seismic activity due to dense fault networks. The strength, location, and timing of earthquakes depend on fault shape and orientation, stress interactions with neighboring faults, rock properties, and many other factors, complicating the task of modeling and predicting events (U.S. Geological Survey [USGS], 2019).

To study these factors, scientists use a variety of measurements. GPS and satellite data track slow tectonic movements and crustal deformation. Seismometers record earthquake waves, providing insight into fault behavior and energy release. Stress and strain in rocks are measured via geological mapping, rock samples, and lab experiments simulating underground pressure. Historical earthquake records provide additional context for modeling seismic conditions. Combining these methods allows researchers to identify relationships between these variables and form basic predictive models. Machine learning builds on these foundations, creating more complex models for analyzing seismic activity.

Seismic events are categorized by magnitude, depth, and focal mechanism. Large earthquakes are rare but impactful, whereas smaller events occur more frequently and can offer precursory signals. Understanding these mechanisms is crucial for ML applications: models must detect patterns in diverse data, capture both rare high-magnitude events and frequent smaller ones, and account for nonlinear spatial and temporal relationships (Asim et al., n.d.).

ML complements traditional seismology by identifying subtle correlations in complex datasets, including seismic waveforms, geodetic measurements, and fault system attributes. Key approaches include supervised learning for labeled data, unsupervised learning for anomaly or cluster detection, and hybrid methods that combine physical models with data-driven techniques. Common tools include decision trees, ensemble methods, statistical models, and deep learning architectures such as convolutional and recurrent neural networks. These provide the foundation for the methodological discussions that follow.

Data Sources and Infrastructure

ML for earthquake prediction depends on large, diverse datasets. In the U.S., the USGS provides historical earthquake records including location, magnitude, depth, and focal

mechanism, while the California Department of Conservation maintains state-level seismic data, fault maps, and regional stress measurements.

Realtime monitoring adds crucial input for ML models. Seismometers record ongoing ground motion, while GPS and satellite measurements track slow tectonic movements and strain along faults. Some models also use remote sensing like InSAR to detect subtle ground deformation.

These datasets are processed and combined into usable formats for ML algorithms. By feeding historical and real-time data into models such as decision trees, neural networks, and deep learning architectures, ML can detect complex patterns, correlations, and precursors in seismic activity that traditional statistical methods often miss. This infrastructure is therefore essential for training, testing, and improving earthquake prediction models.

Main Body: ML Techniques and Applications for Earthquake Prediction

1. Classical ML Models

Classical ML models, such as Decision Trees, Random Forests, and Support Vector Machines, are designed to work with structured data, meaning data that can be represented in tables with clear features like magnitude, depth, location, and frequency of earthquakes. These models learn patterns and relationships between these features to make predictions, for instance, classifying earthquakes into categories or estimating their potential severity. They do not require massive amounts of data or complex computation, and their outputs are easier to interpret than deep learning models.

A concrete example comes from a study by Mohan et al. (n.d.), where Random Forest and SVM models were applied to a global historical earthquake dataset. The study aimed to classify earthquakes as significant (magnitude 6.0 or higher) or minor, analyzing patterns in past seismic events rather than predicting future occurrences. The Random Forest model achieved roughly 88 percent accuracy and a 0.96 area under the curve score, reliably distinguishing strong earthquakes from weaker ones. While these models are limited in predicting exact timing or location and are rarely used in real-time operational systems, the patterns found from their close analysis of historical data serve as the baseline for the development and improvement of other predictive models.

2. Deep Learning for Magnitude Estimation

Deep learning approaches, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs) like LSTMs, are designed to automatically learn complex patterns in data without the need for manually engineered features. These models excel at recognizing sequences and subtle relationships in time-series or waveform data, making them well suited to modeling the continuous signals recorded by seismometers. Deep learning can uncover patterns that classical models might miss, such as the way small precursor vibrations evolve into a larger earthquake.

A study by Zhu, Li, and Song (2025) developed a deep learning model to estimate the magnitude of an earthquake using waveform data from a single station. Convolutional neural network (CNN) layers extracted relevant features from the waveforms, while long short-term memory (LSTM) layers tracked temporal changes in these features. The model achieved estimates with an average error near zero and a standard deviation of approximately 0.2 magnitude units. The models accurately pinpointed magnitudes for ongoing earthquakes within 1-3 seconds after receiving the seismic waveforms, in addition to also performing much more robustly in complex seismic regions. Compared to the techniques used by traditional seismometers from seismic stations, which took up to between 2 - 10 minutes from data acquisition to the end of analysis to provide a stable estimation of magnitude. (US Geological Survey, n.d.). The waveform data and geological patterns captured by these models can feed into other predictive systems, enabling more responsive early warning protocols where seconds can be critical.

3. Hybrid Deep Learning and Forecasting Models

Hybrid models combine the strengths of ML with traditional statistical or physics-based models to improve forecasting. They often integrate data-driven pattern recognition with established empirical laws to enhance accuracy and stability. In earthquake research, hybrid approaches can combine deep learning's ability to capture spatial and temporal correlations with statistical models that account for aftershock rates or fault behaviors.

A notable hybrid example is the CL-ETAS model developed by Zhang, Ke, Liu, and Zhang (2024), which merges a Convolutional LSTM neural network with the Epidemic-Type Aftershock Sequence (ETAS) model. ETAS estimates the probability and expected distribution of aftershocks based on past seismic events, while the ConvLSTM component detects complex spatiotemporal patterns in historical seismic data. Using Southern California earthquake records, the hybrid model outperformed both ConvLSTM and ETAS alone, producing more stable and accurate forecasts of moderate aftershocks and their spatial distribution following major earthquakes (Zhang et al., 2024). However, the model's predictive capacity is primarily limited to aftershock sequences and cannot reliably forecast the occurrence or timing of major earthquakes themselves. Rather, the hybrid approach functions more as a nowcasting tool for aftershocks, providing real-time or near-real-time probabilistic estimates immediately following a major event. Despite these limitations, the model provides practical value for emergency response and hazard mitigation, as aftershocks can themselves inflict substantial damage and hinder recovery efforts, often just as significant as the mainshocks themselves.

4. Temporal and Spatial Deep Learning Models

Temporal and spatial deep learning models, including transformer networks, CNN-RNN combinations, and attention-based architectures, are designed to capture relationships across both space and time. They can detect patterns such as how seismic activity propagates along fault lines over months or years, or how signals at different stations relate to each other. Unlike

classical models, these approaches do not require manual feature engineering and can process raw sensor data directly, which allows them to model complex seismic dependencies that vary over time and location.

For example, a study by Adebbe, Kebede, Kevin, and Demissie (2023) using transformer architectures in the Horn of Africa compared them with LSTM and BiLSTM models. The transformers, including attention layers that focus on important data points, outperformed the other approaches in predicting earthquake magnitudes from seismic waveform and catalog data associated with already detected events, rather than forecasting the occurrence of future earthquakes. The study showed that transformers can learn long-range dependencies, such as temporal and spatial relationships within recorded seismic signals across multiple stations, and can integrate signals from multiple stations to produce more accurate predictions. These models require large datasets and significant computing resources but demonstrate the potential of modern deep learning to analyze complex earthquake patterns in retrospective magnitude estimation tasks where sufficient data exist.

5. Aftershock Forecasting with Machine Learning

Aftershocks are smaller earthquakes that follow a mainshock, and predicting them is important for emergency response and hazard assessment. Machine learning models targeting aftershocks focus on identifying patterns within clusters of seismic events to forecast the likelihood and severity of subsequent quakes. Unlike full earthquake prediction, these models do not try to forecast earthquakes from scratch but provide actionable information on short-term risks after a main event.

The NESTORE algorithm, developed by Eleni-Apostolia Anyfadi and colleagues, was tested on Greek seismic data. Using features like energy release, number of events, and timing, NESTORE classified aftershock clusters and predicted which ones were likely to include strong events. The algorithm correctly identified critical clusters within six hours of the mainshock, showing high accuracy. demonstrates that machine learning can be valuable for monitoring aftershocks and improving hazard response, even if predicting mainshocks remains beyond current capabilities.

Conclusion

Machine learning has significantly advanced earthquake research by enabling more sophisticated analysis of seismic data, improving magnitude estimation, event classification, and aftershock nowcasting. Across classical, deep learning, and hybrid models, ML reveals patterns that traditional methods often miss, providing valuable insights for hazard assessment and real-time situational awareness. Many of the methodologies of ML already have real-world applications in the field, from supporting early warning systems to planning out improved disaster preparedness strategies. However, reliably forecasting the timing and location of major earthquakes remains beyond current capabilities, and most models are limited to retrospective analysis or short-term predictions. How the field should progress as a whole is hard to determine, as the innovations of ML in seismic prediction is a relatively recent development, leaving

substantial room for refinement and optimization in the near future. Overall, ML represents a promising step toward data-driven earthquake prediction, with the potential to transform preparedness and risk mitigation strategies.

Works Cited

- Abebe, E., Kebede, H., Kevin, M., & Demissie, Z. (2023, March 24). Earthquakes magnitude prediction using deep learning for the Horn of Africa. *Soil Dynamics and Earthquake Engineering*, 170, 107913. <https://doi.org/10.1016/j.soildyn.2023.107913>
- Anyfadi, E.-A., Gentili, S., Brondi, P., & Vallianatos, F. (2023). Forecasting Strong Subsequent Earthquakes in Greece with the Machine Learning Algorithm NESTORE. *Entropy*, 25(5), 797. <https://doi.org/10.3390/e25050797>
- Asim, K. M., Idris, A., Iqbal, T., & Martínez-Álvarez, F. (n.d.). Earthquake prediction model using support vector regressor and Hybrid Neural Networks. *PLOS ONE*. <https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0199004>
- Feng, B., & Fox, G. C. (2021, September 11). Spatiotemporal pattern mining for nowcasting extreme earthquakes in Southern California. *arXiv.org*. <https://arxiv.org/abs/2012.14336>
- Fox, G., Rundle, J., Donnellan, A., & Feng, B. (2021, December 18). Earthquake nowcasting with Deep Learning. *arXiv.org*. <https://arxiv.org/abs/2201.01869>
- Michael, A. J., & Langbein, J. O. (1993, March 30). Earthquake prediction lessons from Parkfield experiment. *Eos*, 74(13), 145–155. <https://doi.org/10.1029/93EO00313>
- Mohan, K. A. (2025, July 30). Earthquake Prediction Using Machine Learning Techniques. *IJRASET.com*. <https://www.ijraset.com/research-paper/earthquake-prediction-using-machine-learning-techniques>
- U.S. Geological Survey. (n.d.). How quickly is earthquake information posted to the USGS website and sent out via the Earthquake Notification Service (ENS) and other feeds? U.S. Department of the Interior. <https://www.usgs.gov/faqs/how-quickly-earthquake-information-posted-usgs-website-and-sent-out-earthquake-notification>
- Yavas, C. E., Chen, L., Kadlec, C., & Ji, Y. (2024, October 18). Improving earthquake prediction accuracy in Los Angeles with Machine Learning. *Nature News*. <https://doi.org/10.1038/s41598-024-76483-x>
- Zhang, H., Zhang, Y., Liu, W., & Ke, S. (2024). A combining earthquake forecasting model between deep learning and epidemic-type aftershock sequence (ETAS) model. <https://ouci.dntb.gov.ua/en/works/ldeK5em9>
- Zhu, J., Li, S., & Song, J. (2025, August 25). Multimodal Deep Learning Network for fast seismic discrimination and magnitude classification - *geoscience letters*. SpringerLink. <https://link.springer.com/article/10.1186/s40562-025-00412-7>

Leveraging Existing Nanostore Infrastructure for Inclusive Market Access: Consignment-Based Entrepreneurship in Jordan By Leyth Sharaf

Abstract

Neighborhood corner stores, commonly termed “nanostores” in logistics and retail scholarship remain a prevalent retail format across many developing economies. Despite their ubiquity, nanostores are rarely considered to be institutional platforms for inclusive entrepreneurship. This paper advances the argument that Jordan’s neighborhood grocery shops (bakālāt) can be conceptualized as consignment-based market-access infrastructure, particularly for women, youth, and refugees who face systemic barriers to conventional forms of entrepreneurship.

Drawing on supply-chain and logistics literature, alongside development research on labor-market exclusion in Jordan, the paper proposes a conceptual framework in which home-based or small-scale producers retain ownership of goods until sold, while bakālāt provide shelf space, customer foot traffic, and neighborhood-embedded trust relationships. Such arrangements may reduce entry barriers by lowering upfront capital requirements and easing transportation- and childcare-related costs and access constraints that typically limit market participation. The framework is offered as an analytical tool for examining how informal retail infrastructure may facilitate access to markets for entrepreneurship.

Keywords nanostores; inclusive entrepreneurship; market access; consignment retail; informal economy; women’s economic participation; youth employment; refugees; neighborhood retail; Jordan; bakālāt

1. Introduction

Across emerging and developing economies, very small neighborhood retail outlets, commonly referred to as nanostores, play a central role in everyday consumption and informal economic activity. In the supply-chain literature, nanostores are defined as small, traditional, family-operated outlets characterized by limited physical space, low capital requirements, and minimal reliance on formal logistics or information technology systems (Ortega Jimenez et al., 2023).

These stores are deeply embedded within residential neighborhoods and cater to daily, small-volume demand for fast-moving consumer goods. Their operations rely heavily on informal practices such as cash transactions, verbal agreements, and personalized customer relationships, rather than formal contracts or inventory software. Their continued prevalence reflects their adaptive capacity in environments of fragmented demand, limited purchasing power, and constrained transport infrastructure.

While existing scholarship has generated substantial insights into nanostore operations, replenishment, and last-mile distribution, it has largely overlooked nanostores’ potential role as producer-facing market-access platforms. This omission is especially salient in Jordan, where

dense neighborhood retail networks coexist with persistently low labor-market participation among women, youth, and refugees. Addressing this gap, the paper positions bakālāt as consignment-based market-access infrastructure and develops a theoretically grounded pathway through which small-scale producers may connect to local demand via existing neighborhood retail systems.

2. Literature Review

2.1 Nanostores in Logistics and Supply-Chain Research

The term “nanostore” was formally introduced around 2013 to describe small-scale, family-run, proximity-based retail shops prevalent across the Global South (Al-Talib et al., 2020). These stores often operate in spaces ranging from 15 to 40 square meters and are typically embedded in residential neighborhoods.

Ortega Jimenez et al. (2023) offer a comprehensive examination of the nanostore literature, drawing on more than 430 peer-reviewed publications produced between 2014 and 2023. Their review estimates that approximately 50 million nanostores operate across developing countries, accounting for around half of total retail activity in many national contexts. In Latin America, nanostores are estimated to represent between 60 and 70 percent of the consumer-packaged goods market (Ortega Jimenez et al., 2023).

Research in this domain primarily focuses on operational aspects such as last-mile delivery, replenishment strategies, retailer–supplier relationships, and digital tool adoption. The high spatial concentration of nanostores creates logistics challenges, including low drop-size deliveries, high delivery frequency, and intensive cash handling, rendering them distinct from formal supermarket formats.

2.2 Informality, embedded relationships, and retail governance

Beyond their retail function, nanostores function as socially embedded economic institutions, rather than merely transactional outlets. Boulaksil et al. (2019) document the widespread provision of informal, interest-free credit to known customers, with repayment obligations enforced through interpersonal trust rather than legal contracts. These trust-based arrangements allow households to smooth consumption amid income volatility and limited access to formal financial services.

Importantly, Boulaksil et al. (2019) also show that informality extends beyond customer-facing transactions to shape supply-chain coordination itself. Their empirical study of 333 nanostores across eight Moroccan cities finds that ordering behavior is unstructured, irregular, and largely reactive, often based on immediate cash availability, perceived stockouts, or habitual routines, rather than formal forecasting or inventory systems. This lack of formal coordination reflects not inefficiency per se, but a governance logic shaped by cash-flow limitations and relationship-based practices. Nanostores rely heavily on embedded social relationships to manage access to goods, liquidity, and information at the neighborhood

level. From a logistics perspective, nanostores also function as key demand points within local supply chains. However, because ordering decisions are often driven by short-term cash availability rather than anticipated customer demand, retailer orders may not accurately reflect underlying consumption patterns. As a result, suppliers and distributors responsible for delivering goods to neighborhood retailers face coordination challenges, including frequent small deliveries, underutilized transport capacity, and inefficient routing.

Furthermore, Boulaksil et al. (2019) identify heterogeneity in retail governance structures, showing that nanostores vary depending on ownership model (owner-run, family-run, employee-managed) and operator characteristics, such as education level. These differences affect ordering frequency, credit provision, and sales volumes, resulting in uneven transportation demand across stores.

These findings resonate with broader scholarship on the informal economy, which challenges the notion that informality is a transitional or marginal phenomenon. In particular, Chen (2012) argues that informality is not a residual category awaiting formalisation, but rather a pervasive and structurally embedded mode of economic organisation in most developing economies. Drawing on the framework developed by the International Labour Organization and the Women in Informal Employment: Globalizing and Organizing (WIEGO) network, Chen distinguishes between informal enterprises and informal employment, both of which are governed by their own internal rationalities, institutional logics, and survival strategies.

Chen asserts that informal systems evolve as rational responses to institutional exclusion including weak legal enforcement, barriers to formal credit, and inappropriate or unaffordable regulatory regimes. From this perspective, informal arrangements such as cash-based transactions, trust-based enforcement, and flexible labor arrangements are not simply coping mechanisms, but functional governance adaptations that offer efficiency and reliability under conditions of structural constraint. Informal actors are therefore not outside the economy, but constitutive of it, operating within alternative institutional arrangements that are responsive to exclusion, ambiguity, and volatility.

In the context of nanostores, this framing invites a rethinking of operational “inefficiencies” such as irregular ordering or liquidity-driven stock decisions. These behaviors are not symptoms of disorganization, but manifestations of an alternative economic rationality that prioritizes resilience, relational trust, and immediacy of cash flow over long-term optimization. Nanostores thus exemplify what Chen refers to as “institutional informality,” a governance system in which market coordination is achieved through relational and reputational mechanisms rather than through codified contracts or digital inventory systems.

This perspective positions nanostores as critical infrastructure in their own right, not merely as imperfect versions of formal retail. Their decision-making, shaped by liquidity, trust, and embeddedness, contributes directly to the structure of last-mile supply chains. Consequently, coordination frictions observed in nanostore environments should be interpreted not only as logistical challenges, but as inter-system mismatches between informal governance at the retail level and formal, data-driven planning systems at the distributor or manufacturer level.

2.3 Market Exclusion and Entrepreneurship in Jordan

Development research consistently documents deep structural barriers to labor-market participation in Jordan, particularly among women, youth, and refugees. While educational attainment has risen steadily, this has not translated into commensurate improvements in employment or entrepreneurship outcomes, especially for women. According to the Organisation for Economic Co-operation and Development (OECD), the International Labour Organization (ILO), and the Center of Arab Women for Training and Research (CAWTAR), Jordan ranks among the lowest in female labor-force participation across the MENA region, despite decades of investment in women's education (OECD, ILO, & CAWTAR, 2020). This disconnect reflects a set of structural barriers that impede entry into formal and semi-formal economic activity, particularly for women, youth, and displaced populations.

One of the most persistent barriers to entrepreneurship is access to capital, particularly for women and youth. Women in Jordan face compounded financial exclusion arising from limited asset ownership, discriminatory inheritance practices, and lack of collateral, all of which hamper access to formal credit markets. These impediments are not merely procedural but are deeply embedded in legal structures and customary norms. For instance, women in the MENA region face among the highest levels of discrimination globally in access to productive and financial resources. The Social Institutions and Gender Index estimates discrimination levels of approximately 45 percent in Northern Africa, compared with a global average of 27 percent. Ownership of land and non-land assets also plays a central role in entrepreneurial capacity, serving both as a productive input and as collateral for financing. While inheritance rights are formally codified under Sharia-based personal status law, women typically receive smaller inheritance shares than men and often face additional social pressures that limit their effective control over inherited assets (World Bank, 2019; OECD, ILO, & CAWTAR, 2020). These dynamics restrict women's ability to accumulate wealth or leverage assets for entrepreneurial activity. As a result, even modest startup costs such as inventory purchase, shop rental, or licensing fees can present prohibitive barriers. Because assets function simultaneously as inputs and as collateral, unequal ownership directly translates into exclusion from formal finance, particularly for young women and refugees lacking employment histories or credit records.

This unequal distribution of productive assets significantly raises the fixed costs of entry into entrepreneurship. Even small-scale retail initiatives such as renting a storefront or purchasing initial inventory become inaccessible without borrowing, which is in turn unavailable without assets. For youth, the challenge is compounded by absence of credit history, limited savings, and low initial earning potential.

The unequal distribution of unpaid care responsibilities is another critical factor limiting women's market participation. Time-use surveys show that women devote significantly more hours to caregiving, housework, and eldercare, which restricts their availability for time-intensive entrepreneurial activity or jobs requiring long hours away from home (OECD, ILO, and CAWTAR, 2020). These responsibilities interact with mobility restrictions, including

gendered safety concerns, cultural expectations, and weak public transportation, making it difficult for women to operate businesses in commercial districts or distant markets.

The OECD, drawing on ILO data, documents that a large share of women's employment in Jordan occurs in informal or vulnerable categories, including own-account work, contributing family labor, and home-based activities (OECD, ILO, & CAWTAR, 2020). While such arrangements enable participation under restrictive conditions, they are typically characterized by low productivity, income volatility, and limited access to social protection. Informality thus functions simultaneously as a coping mechanism and a structural constraint, permitting economic activity while limiting opportunities for growth, market access, and income stability.

This exclusion from downstream market linkages is particularly consequential in retail-oriented sectors, where access to shelf space, customer footfall, and trusted intermediaries is critical. This lack of "market access" is consistent with broader development findings that small producers and micro-entrepreneurs often face binding barriers not primarily in skills or willingness to work, but in reliably connecting to demand channels and distribution networks. The World Development Report 2020 argues that participation in value chains and the gains from trade depend heavily on the ability of smaller actors to access markets and link to buyers, intermediaries, and distribution systems (World Bank, 2020). The OECD notes that credibility, reputation, and family standing remain decisive factors shaping women's participation in commerce (OECD, ILO, & CAWTAR, 2020).

While institutional analyses document structural barriers to market participation, public-opinion data provide insight into how these obstacles are experienced at the household level. Nationally representative survey evidence from the Arab Barometer's Wave VIII Jordan Public Opinion Report indicates that economic conditions and job availability remain the most salient concerns among Jordanian citizens. A majority of respondents identify the economy and employment opportunities as the country's most pressing challenges, reflecting persistent perceptions of economic insecurity and limited pathways into stable income generation (Arab Barometer, 2024). The Arab Barometer findings further suggest that formal employment is not widely perceived as a reliable or accessible route to economic stability, particularly among younger cohorts. These perceptions help explain the continued reliance on informal, supplementary, or neighborhood-based income-generating activities, even among individuals with educational qualifications. In this context, economic participation is often pursued through locally embedded arrangements rather than through formal labor-market integration alone (Arab Barometer, 2024).

Importantly, these barriers do not operate in isolation but reinforce one another. For example, a woman lacking collateral may also face mobility limits due to childcare duties, and restricted market access due to limited social capital. For youth, the absence of work experience compounds asset and network exclusion. Refugees, especially women, may simultaneously face all of the above, capital exclusion, legal barriers, social marginalization, and logistical constraints, rendering traditional entrepreneurship pathways nearly inaccessible.

The conditions outlined above indicate that barriers to economic participation in Jordan extend beyond deficits in skills or motivation. Instead, they point to a market-access problem in which productive capacity may exist, but entry into reliable channels of demand remains difficult on economically viable and socially acceptable terms. This interpretation aligns with the World Bank's Jobs and Economic Transformation (JET) framework, which emphasizes "creating and connecting to markets" as central to improving livelihood outcomes, particularly in contexts where formal employment creation is insufficient to absorb labor supply (World Bank, 2019). From this perspective, access to proximate and repeated demand interfaces plays a critical role in shaping feasible pathways into economic participation.

This intersectionality highlights why entrepreneurship models premised on independent storefronts, full-time market presence, or significant startup capital are often infeasible. It also underscores the relevance of alternative models, such as consignment-based partnerships with embedded retailers, which reduce risk exposure, eliminate capital barriers, and localize market access within trusted neighborhood networks.

3. Discussion

Jordan's retail landscape is characterized by a dense network of small, neighborhood-based grocery and convenience outlets that operate as part of the informal or semi-formal economy. These *bakālāt* are embedded within daily consumption patterns and rely on proximity, reputation, and repeat interactions rather than scale, branding, or formal logistics systems.

Within this setting, consignment-based retail arrangements emerge as a theoretically plausible mechanism for lowering entry barriers to market participation. Under consignment, producers retain ownership of goods until sale, while shopkeepers provide shelf space and customer access in exchange for an agreed revenue share. This model reduces the need for upfront capital investment, inventory financing, and frequent physical presence at the point of sale. The framework therefore positions nanostores not as agents of transformation in themselves, but as enabling nodes within local economic networks. By leveraging infrastructure that already exists, physical space, customer foot traffic, and embedded trust, consignment partnerships can potentially convert latent productive capacity into market-facing economic activity. This approach differs from entrepreneurship models that require formal registration, external financing, or relocation to commercial districts, all of which can constitute significant barriers for marginalized groups.

Importantly, this framework does not assume that nanostores automatically generate inclusive outcomes. Rather, it advances a capability-conversion logic, whereby nanostores possess certain structural characteristics (proximity, trust, and already incurred fixed retail costs) that *can* be activated through institutional arrangements such as consignment agreements, basic record-keeping, or light-touch digital tools. Whether these characteristics translate into inclusive economic participation depends on governance, incentives, and coordination mechanisms.

By situating nanostores within a broader ecosystem of producers, consumers, and intermediaries, the framework aligns with recent supply-chain research that emphasizes the role of small retail actors as flexible nodes capable of absorbing innovation when supported by appropriate coordination mechanisms.

Figure 1. Conceptual Framework: Nanostores as Consignment-Based Market-Access Platforms

Home-based or small-scale producers

(women, youth, refugees)



Consignment arrangement

– producer retains ownership

– no upfront inventory purchase



Neighborhood nanostores (bakālāt)

– existing retail space

– trust-based relationships

– local customer access



Local consumers

– proximity-driven demand

– repeat purchasing



Outcomes (conceptual)

– reduced entry barriers

– income generation without storefront ownership

– enhanced local economic participation

Note: The figure represents a conceptual mechanism rather than an empirical claim. Outcomes depend on institutional design, incentives, and contextual factors.

For individuals such as women, youth, or refugees who face barriers related to childcare responsibilities, mobility limitations, or legal and financial exclusion, consignment arrangements may offer a means of participating in commerce without assuming the full risks and costs of operating an independent storefront. By decoupling production from point-of-sale ownership, consignment models alter the distribution of risk and responsibility in ways that may expand feasible participation at the margins.

Importantly, the potential relevance of bakālāt lies not in their capacity to substitute for formal enterprises or employment, but in their role as institutional complements within Jordan's existing economic landscape. The nanostore literature cautions that inclusive outcomes are not automatic and depend on governance arrangements, incentive alignment, and basic coordination

mechanisms, including clarity around pricing, settlement, and record-keeping. Without such safeguards, market-access initiatives risk reproducing informality without generating sustainable income opportunities.

4. Conclusion

This paper underscores two implications for Jordan. First, policies and programs aimed at expanding entrepreneurship among women, youth, and refugees may benefit from engaging with existing micro-retail infrastructure rather than attempting to bypass it. Second, framing *bakālāt* as potential market-access platforms shifts the focus of intervention from individual-level capacity building alone toward the design of intermediary arrangements that lower entry thresholds into local markets. The effectiveness of such approaches, however, remains an empirical question and depends on how consignment models are structured, regulated, and supported in practice.

Proximity-based market access reduces transport costs and logistical complexity, while embedding economic activity within familiar community settings may lower social and informational barriers. Policies that recognize neighborhood retail spaces as legitimate sites of economic participation could therefore broaden the range of viable livelihood strategies without requiring large-scale infrastructural investment. This neighborhood-channel logic is also consistent with World Bank arguments that participation and upgrading depend on connectivity to demand and buyers; when mobility and time constraints restrict participation in distant commercial districts, locally embedded channels can function as pragmatic “market connections” that expand feasible economic participation at the margin.

Consignment models are particularly relevant in contexts where access to startup capital is limited. By shifting inventory risk away from producers and allowing payment upon sale, consignment arrangements reduce the need for upfront investment and may therefore widen participation among those otherwise excluded from market entry. A central implication is that entrepreneurship-support initiatives need not rely exclusively on the creation of new enterprises or formal retail spaces. Instead, they can leverage the already dense and socially embedded network of neighborhood grocery stores. In Jordan, where *bakālāt* are woven into everyday consumption practices, this suggests scope for programs that position them as intermediaries between small-scale producers and local consumers.

From a policy perspective, this approach aligns with a broader shift away from enterprise creation as a binary formal–informal distinction, toward recognising intermediate market arrangements that lower entry barriers while maintaining economic viability. Supporting consignment-based partnerships through pilot programs, standardised contracts, or light-touch regulation could allow such models to operate transparently without imposing compliance costs that would discourage participation. At minimum, program design would require clear terms for product acceptance, shelf-time expectations, revenue splits, and return or write-off rules to reduce ambiguity for both producers and shopkeepers.

Program design in this area could focus on risk-sharing mechanisms, such as partial guarantees, revolving funds, or partnerships with NGOs and development actors, rather than

direct lending. Importantly, such measures would complement rather than replace existing financial-inclusion efforts, recognising that credit access alone does not resolve market-access barriers. In addition, because nanostores often operate with cash-based settlement practices, transparent and low-burden settlement procedures are essential to prevent payment delays or disputes.

However, these benefits are likely to depend on whether participation can occur without imposing additional mobility, documentation, or compliance burdens that would exclude precisely the target groups. In the Jordanian context, this suggests a role for municipalities, NGOs, and development agencies in facilitating simple contractual standards, transparent revenue-sharing arrangements, and monitoring mechanisms that protect both producers and shopkeepers. Such safeguards are particularly important when working with vulnerable populations, including refugees, to ensure that consignment arrangements do not reproduce exploitative dynamics under the guise of inclusion. Framing *bakālāt* as gateways to entrepreneurship must therefore be accompanied by attention to governance, accountability, and fair commercial practices. Because nanostores commonly rely on informal, relationship-based practices, including informal credit in some settings, program safeguards should prioritise transparency and documentation while remaining light-touch enough to be feasible for small retailers.

Importantly, this does not imply that consignment-based participation is a comprehensive solution to labor-market exclusion; rather, it positions such arrangements as institutional complements that can lower initial thresholds for market participation while broader reforms remain necessary.

These implications point toward a reframing of inclusive entrepreneurship policy in Jordan. Rather than focusing exclusively on individual entrepreneurs, interventions could target market intermediaries and local retail ecosystems as sites where inclusion can be operationalised. By lowering capital, mobility, and market-access barriers simultaneously, nanostore-based consignment models may complement existing employment and small and medium-sized enterprise strategies, particularly in neighborhoods where formal job creation remains limited.

At the same time, the analysis underscores that such models are not a substitute for broader structural reforms related to labor markets, childcare provision, or legal status for refugees. Instead, they represent a pragmatic, incremental approach that builds on existing economic structures while expanding participation at the margins.

Works Cited

- Al-Talib, M., et al. "Achieving Resilience in the Supply Chain by Applying IoT Technology." *Procedia CIRP*, vol. 91, 2020, pp. 752–757, <https://doi.org/10.1016/j.procir.2020.02.231>
- Arab Barometer. Jordan Public Opinion Report: Wave VIII. Arab Barometer, 2024, https://www.arabbarometer.org/wp-content/uploads/ABVIII_Jordan_Report-Public-Opinions-EN-1.pdf
- Boulaksil, Youssef, et al. "Understanding the Fragmented Demand for Transportation in Urban Areas: Evidence from Nanostores in Morocco." *Transportation Research Part A: Policy and Practice*, vol. 129, 2019, pp. 49–63, <https://doi.org/10.1016/j.tra.2019.09.003>.
- Chen, Martha A. *The Informal Economy: Definitions, Theories and Policies*. WIEGO Working Paper No. 1, Women in Informal Employment: Globalizing and Organizing (WIEGO), 2012, https://www.wiego.org/wp-content/uploads/2019/09/Chen_WIEGO_WP1.pdf
- Fransoo, Jan C., et al., editors. *Reaching 50 Million Nanostores: Retail Distribution in Emerging Megacities*. CreateSpace Independent Publishing Platform, 2017.
- Kelley, Donna, et al. *Global Entrepreneurship Monitor 2016/2017 Report on Women's Entrepreneurship*. Global Entrepreneurship Research Association (GERA), Babson College and Smith College, 2017, <https://www.gemconsortium.org/report/gem-20162017-womens-entrepreneurship-report>
- OECD, International Labour Organization, and CAWTAR. *Changing Laws and Breaking Barriers for Women's Economic Empowerment in Egypt, Jordan, Morocco and Tunisia*. OECD Publishing, 2020, https://www.oecd.org/en/publications/changing-laws-and-breaking-barriers-for-women-s-economic-empowerment-in-egypt-jordan-morocco-and-tunisia_ac780735-en
- Ortega Jimenez, Cesar H., et al. "Logistics and Information Technology: A Systematic Literature Review of Nanostores from 2014 to 2023." *Proceedings of the 21st LACCEI International Multi-Conference for Engineering, Education and Technology*, 2023, https://laccei.org/LACCEI2023-BuenosAires/papers/Contribution_726_a.pdf
- Rentech Digital. *List of Grocery Stores in Jordan*. Rentech Digital, 2025, <https://rentechdigital.com/smartscraper/business-report-details/list-of-grocery-stores-in-jordan>
- World Bank. *IDA19: An Overview—Ten Years to 2030: Growth, People, Resilience*. World Bank Group, 2019, <https://documents1.worldbank.org/curated/en/898721564074799426/pdf/IDA19-An-Overview-Ten-Years-to-2030-Growth-People-Resilience.pdf>
- World Bank. *World Development Report 2020: Trading for Development in the Age of Global Value Chains*. World Bank, 2020, <https://www.worldbank.org/en/publication/wdr2020>

The Chemical and Energetic Constraints of Chemosynthesis-Based Ecosystems in Sunless Environments By Neeva Gangopadhyay

Abstract

Most ecosystems on Earth are sustained by photosynthesis, a metabolic process that relies on sunlight to produce organic compounds. However, organisms in sunless environments obtain energy through chemosynthesis, which uses chemical reactions to fix inorganic carbon into organic matter. Chemoautotrophic extremophiles thrive in environments once thought to be uninhabitable, such as deep-sea hydrothermal vents, subsurface rock formations, and hot springs. This literature review examines the primary chemical and energetic limitations of chemosynthesis-based ecosystems and evaluates how these constraints influence the potential for life in extraterrestrial environments lacking sunlight. Evidence from Earth's extremophiles suggests that while chemosynthesis can sustain life under extreme conditions, limited energy availability and chemical constraints restrict ecosystem size and complexity. When applied to environments such as Europa and Enceladus, these limitations indicate that microbial life may be possible, but the development of large, diverse ecosystems is unlikely.

Research Background

Photosynthesis is the primary metabolic process on Earth, powering most ecosystems. Plants, algae, and some bacteria capture energy from the sun to produce oxygen and glucose. Other organisms in the ecosystem consume the glucose, herbivores obtain energy by eating plants, and carnivores obtain energy by consuming herbivores. Photosynthesis is considered the basis of life because it sustains most organisms on Earth. But the discovery of a deep-sea hydrothermal vent in 1977 challenged the assumption that life was only possible with sunlight. An ecosystem containing clams and tube worms was discovered near the vent. Communities of microbes fed off chemicals coming from the vent, converting carbon dioxide into organic compounds. Ecosystems that don't have access to sunlight use chemosynthesis, a metabolic pathway using chemical energy to create food. Some habitats using chemosynthesis include hydrothermal vents, cold seeps, and sulfuric hot springs. The habitats of extremophiles have shaped microbial communities and their cellular mechanisms. Each habitat has an extreme chemical or physical parameter that shows how the organism adapted to its environment over time. Some of these conditions include high or low pressure and low water activity.

Chemosynthesis and the Habitats of Extremophiles

Chemosynthesis is the process by which carbon molecules, typically carbon dioxide or methane, and nutrients are converted into organic matter. The metabolic process uses the oxidation of chemical compounds lacking carbon-hydrogen bonds, such as hydrogen gas or hydrogen sulfide, as a source of energy. There isn't one pathway for chemosynthesis; different chemosynthetic microbe species employ different pathways to harness energy from chemicals. For example, an extremophile employing chemosynthesis is a thermophile. They live around

hydrothermal vents and use sulfur-based chemosynthesis. Waters infused with chemicals emerge from under the seafloor through hydrothermal vents. Some of the chemical reactions release chemical energy, so the microbes use the energy released to convert inorganic carbon to organic molecules through carbon fixation. Carbon fixation is the process by which living organisms convert inorganic carbon, such as carbon dioxide, into organic compounds like sugar. Chemosynthesis supports ecosystems by providing the energy and biomass for life in environments without sunlight.

Most thermophilic organisms live in geothermal environments such as hot springs. Hot springs are beneath the Earth's surface, containing hot groundwater that erupts. The water contains minerals that help these organisms progress. Thermophiles possess a unique protein and lipid structure that enables them to remain stable at these higher temperatures. A specific thermophile employing chemosynthesis is the *Candidatus Desulforudis Audaxviator*. This bacterium lives in groundwater and was discovered 2.8 kilometers below the Earth's surface in a South African gold mine. It lives in extreme conditions: complete darkness, no oxygen, high temperature around 60 degrees Celsius, and lives in a single-species ecosystem. *D. audaxviator* obtains its energy from hydrogen and sulfate produced by the radioactive decay of uranium. The organism also employs a unique form of chemosynthesis. It derives energy from water radiolysis, which splits water into hydrogen and peroxide with energy from radioactive decay found in the surrounding rocks of its habitat. Then, it uses the hydrogen to reduce sulfate to sulfide and fixate carbon dioxide. This creates food for the bacterium and supports an ecosystem from radioactive energy.

Extraterrestrial Environments with Chemical Energy

Some planets have environments similar to those in which extremophiles live. For example, Europa, one of Jupiter's moons, has the parameters necessary for life to live there. Europa is theorized to have a liquid ocean underneath its surface because the moon is crater-free. The planet is very far from the sun and has a surface temperature of around -160 degrees Celsius. But the gravitational push and pull of Jupiter's gravity create friction that warms it, a process called tidal heating. Tidal heating is crucial for life as it supports geological activity, such as volcanic activity and plate tectonics. These processes are essential for nutrient cycling and maintaining a steady temperature. In addition, Europa's volcanic activity could drive hydrothermal circulation, creating an environment similar to thermophiles. Europa is hypothesized to have all the chemicals and energy necessary for life from Jupiter's radiation, similar to the environment of *Candidatus Desulforudis Audaxviator*. Some studies theorize that the effect of radiation on the surface ice of the planet suggests that some of it may break down and produce oxygen and seep through the surface to the ocean below.

One of Saturn's moons, Enceladus, contains the organic compounds necessary for a broader form of chemosynthesis, methanogenesis. Similar to Europa, Enceladus is very cold and contains an ocean below its surface, but it sprays the ocean into space, allowing spacecraft to sample the water. NASA's Cassini spacecraft discovered hydrogen cyanide, one of the most

versatile molecules needed to form amino acids. From the data collected, the compounds in the ocean, carbon dioxide, methane, and hydrogen, indicate that organisms could employ methanogenesis to live. Organisms that use methanogenesis live in marine environments and hydrothermal vents on Earth, although most methanogens are Archaea. Methanogenesis requires no oxygen; the most common pathway involves using hydrogen to reduce carbon dioxide into methane. A larger moon of Saturn also causes tidal heating within the planet, creating an environment similar to that of extremophiles.

Differences between the environments and limitations

Although Earth's chemosynthetic extremophiles demonstrate that life can persist in the absence of sunlight, important chemical and energetic limitations constrain the size, complexity, and longevity of these ecosystems. Chemosynthesis relies on the oxidation of inorganic chemical compounds, which typically generates less energy than photosynthesis. As a result, chemosynthesis-based environments are often dominated by microbial communities rather than complex multicellular organisms. On Earth, chemosynthetic ecosystems such as hydrothermal vents are sustained by continuous geological processes that supply electron donors and acceptors. However, these resources are often restricted and chemically limited, meaning life can only exist when favourable redox reactions are possible. For example, the *Candidatus Desulfurudis audaxviator* illustrates these limitations. It survives in a highly specialized single-species environment supported by hydrogen and sulfate from radioactive decay. While this demonstrates the resilience of life, it also highlights the narrow conditions required for survival.

While these constraints can be applied to extraterrestrial environments like Enceladus and Europa, additional uncertainties arise. Although both of the moons possess subsurface oceans and potential chemical energy sources driven by tidal heating and radiation, the stability of these energy sources over geological timescales is unknown. It is also unclear whether sufficient electron acceptors are available to sustain long-term metabolic processes. Without efficient nutrient cycling and sustained chemical gradients, chemosynthetic life would be restricted to small microbial communities. For example, methanogens are mostly only Archaea, single-cell prokaryotic organisms.

Overall, these limitations suggest that while chemosynthesis increases the range of environments capable of supporting life, it does not eliminate fundamental energetic constraints. In sunless extraterrestrial environments, chemosynthesis may allow for the existence of microbial life, but the emergence of complex ecosystems comparable to those powered by photosynthesis on Earth is unlikely.

Works Cited

- Beall, Abigail. "Mission to Europa." *New Scientist*, 24 Aug. 2024.
- Bhagwan Rekadwad, et al. "Extremophiles: The Species That Evolve and Survive under Hostile Conditions." *3 Biotech*, vol. 13, no. 9, 25 Aug. 2023, doi:10.1007/s13205-023-03733-6.
- Elizabeth, et al. "Editorial: Acidophile Microbiology: From Extreme Environments to Biotechnological Applications." *Frontiers in Microbiology*, vol. 15, 4 July 2024, doi:10.3389/fmicb.2024.1454559.
- Lubofsky, Evan. "The Discovery of Hydrothermal Vents." *Oceanus*, Woods Hole Oceanographic Institution, 11 June 2018, www.whoi.edu/oceanus/feature/the-discovery-of-hydrothermal-vents/.
- Lyu, Zhe, et al. "Methanogenesis." *Current Biology*, vol. 28, no. 13, 9 July 2018, pp. R727–R732, doi:10.1016/j.cub.2018.05.021.
- NASA. "NASA Study Finds Life-Sparking Energy Source and Molecule at Enceladus." NASA, 14 Dec. 2023, www.nasa.gov/missions/cassini/nasa-study-finds-life-sparking-energy-source-and-molecule-at-enceladus/.
- NASA. "Enceladus." *NASA Science*, 4 Nov. 2024, science.nasa.gov/saturn/moons/enceladus/.
- NOAA. "What Is an Extremophile?" NOAA Ocean Service, 2019, oceanservice.noaa.gov/facts/extremophile.html.
- O'Callaghan, Jonathan. "Life on Mars Could Be Surviving in an Area Deep Underground." *New Scientist*, 7 Dec. 2024. Science Reference Source.
- Paulpreuss. "Bold Traveler's Journey toward the Center of the Earth." Berkeley Lab News Center, Lawrence Berkeley National Laboratory, 9 Oct. 2008, newscenter.lbl.gov/2008/10/09/bold-traveler/.
- Ricci, Francesco, and Chris Greening. "Chemosynthesis: A Neglected Foundation of Marine Ecology and Biogeochemistry." *Trends in Microbiology*, 1 Jan. 2024, doi:10.1016/j.tim.2023.11.013.
- "Some Like It Hot! Studying Thermophiles in Yellowstone National Park." U.S. Geological Survey, 4 Sept. 2023, www.usgs.gov/observatories/yvo/news/some-it-hot-studying-thermophiles-yellowstone-national-park.

Substance Abuse and Homelessness, Which Comes First? Is substance abuse the cause of homelessness, or is it the consequence?

By Allison Lu

Introduction

Michael was in his 20s when he found himself jobless and homeless. He recounts in an interview with Spring Rescue Mission:

“I...put on a backpack and just started roaming the streets. I was in a dark place. And then came the meth” (Spring Rescue Mission).

Michael's story portrays a common misconception—that substance abuse is the leading cause of homelessness. While addiction certainly plays a role in an individual's susceptibility to homelessness, it remains an extreme oversimplification for the far more complex reality. The media often perpetuates a one-sided notion that addiction is the primary culprit. However, many studies have emphasized the opposite—that substance abuse often occurs *after* one's homelessness onset, not before.

This blog aims to explore the relationship between substance abuse and homelessness. More specifically, which comes first?

Background

In 2024, there were about 771,400 homeless individuals across the country, with that number only increasing each year, according to the Federal Reserve Bank of Minneapolis (Camner McKay and Cowles). Among these homeless individuals are veterans, families displaced by foreclosure, youth aging out of foster care, and, prominently, individuals struggling with substance abuse. The National Coalition for Homelessness writes that 55% of homeless individuals are alcohol dependent, with 25% abusing other substances. This percentage of substance use is found to be 2-3 times greater than in the housed community (National Coalition for the Homelessness). Although substance abuse may, in many cases, contribute to housing instability, it has also been found to be a consequence rather than a cause of homelessness.

Homelessness leading to substance abuse

Indeed, a growing body of research challenges the notion that addiction is always the inciting factor. One study by Gomez et al. examines how homelessness itself fosters conditions which may lead to substance abuse (Gomez et al.).

Social Support

During an individual's early onset of homelessness, they are often left with little support due to separation from family and social institutions (ie. work). During such times, many homeless individuals turn to “street families” as their primary source of social support. Although

these street families can provide emotional support, safety, and sometimes resources, they are also often influential and risky. With regards to substance use, homeless individuals who found themselves in non-using peer groups were significantly less likely to develop alcohol dependencies. In contrast, those rooted in peer groups where drug use was common were more likely to become substance dependent. In such drug-using homeless communities, abusing alcohol and drugs becomes ideal due to various reasons:

1. Substances provide a daily activity for these peer groups to connect and bond.
2. Using substances is often seen as necessary to fit into the community.
3. Peer groups normalize drug-use and support unhealthy behaviors, which many homeless individuals seek out.

Economic Resources

Beyond social influences, economic factors also play a large role in a homeless individual's substance use patterns. Studies have shown that one's source of income plays a large role in the likelihood of their alcohol dependence. For instance, those who made income through selling blood/plasma were much less likely to abuse alcohol, likely due to the notion of maintaining "clean" blood. This trend remains when looking at drug abuse. Young adults who made their income selling self-made items were much less likely to be drug dependent than those who reported dealing drugs as their main source of income. Researchers have estimated that, in the homeless population, unemployment rates range from 57% to over 90%. Thus, limited economic opportunities while homeless may push individuals towards higher-risk activities which allows them to be more susceptible to substance abuse.

Coping Mechanism

Finally, for many homeless individuals, substances are used as means of coping. Not only do many substances release temporary dopamine, offering an emotional release from the pressures of homelessness, but they also help to address basic survival needs. Substances may allow homeless individuals to stay warm, suppress their hunger, or stay awake in environments where extra precaution is necessary. In this way, substances are used as self-medication and means to help them cope with the reality in which they reside.

Still, the relationship between substance abuse and homelessness is not one-directional.

Substance abuse leading to homelessness

While many individuals begin using substances succeeding their homelessness onset, for others, addiction plays a large role in their loss of housing in the first place. The following statistics suggests this:

- In 2014, the United States Conference of Mayors released a survey revealing that substance abuse was tied for the third leading cause of homelessness in 25 cities surveyed (U.S. Conference of Mayors).
- In multiple studies, more than two-thirds of homeless individuals have described substance abuse as a major cause of their homelessness (AdCare; Spinner and Leaf).
- According to a study conducted by Spinner and Leaf, 25 percent of the homeless individuals surveyed linked drug use as their primary cause of homelessness (Spinner and Leaf).

In many cases, the damage substance abuse is gradual and cumulative. Addiction can lead to worsened relationships with family and friends, unemployment from one's primary source of income, and, eventually, loss in housing. This economic decline, combined with an individual's strained relationships with their support networks (such as their family), can make it extremely difficult for individuals to regain economic and emotional stability, ultimately pushing many into homelessness.

Conclusion

So, what comes first, substance use or homelessness? Ultimately, the answer to this question lies purely on an individualized basis. For some, substances may trigger a chain reaction where the final link leads to an individual's loss of housing. For others, substance use only emerges as means of fitting into street culture. However, what remains consistent is the cyclical nature of the two, each one reinforcing the other, and, in both cases, one condition often worsens the other. Substance use makes it more difficult for homeless individuals to receive housing and aid while homelessness increases an individual's susceptibility to addiction. Without the proper support, individuals may lack the necessary means to escape the endless loop. Further exacerbating the issue, housing assistance programs place many restrictions, hindering those with a history of substance abuse from receiving proper aid. Specifically, applicants must be denied if they...

1. Have been evicted from public housing within the last 3 years due to drug related reasons.
2. Have been convicted of manufacturing methamphetamines on public housing.
3. Are using illegal drugs currently.
4. Are abusing alcohol in a manner that interferes with the public housing community (U.S. Department of Housing and Urban Development).

So, instead of focusing on which comes first, the better question may be: how may we break the cycle of substance abuse and homelessness, regardless of where it begins?

Works Cited

- AdCare Educational Institute. Substance Abuse Among the Homeless Population. AdCare, <https://adcare.com/addiction-demographics/homeless-population/>. Accessed 26 Jan. 2026.
- Camner McKay, Lisa, and Kenneth Cowles. "Who Is Homeless in the United States? A 2025 Update." Federal Reserve Bank of Minneapolis, 13 Mar. 2025, <https://www.minneapolisfed.org/article/2025/who-is-homeless-in-the-united-states-a-2025-update>. Accessed 26 Jan. 2026.
- Gomez, Rebecca, et al. "Factors Associated with Substance Use among Homeless Young Adults." Substance Abuse, vol. 31, no. 1, Jan. 2010, pp. 24–34. PubMed Central, doi:10.1080/08897070903442566.
- National Coalition for the Homeless. Substance Abuse and Homelessness. June 2017, <https://nationalhomeless.org/wp-content/uploads/2017/06/Substance-Abuse-and-Homelessness.pdf>. Accessed 26 Jan. 2026.
- Spinner, George F., and Philip J. Leaf. "Homelessness and Drug Abuse." Psychiatric Services, vol. 43, no. 2, 1992, pp. 166–170, <https://psychiatryonline.org/doi/abs/10.1176/ps.43.2.166>.
- Spring Rescue Mission. "I Threw Up My Hands: Michael's Story of Overcoming Addiction and Homelessness." Spring Rescue Mission, <https://www.springsrescuemission.org/articles/i-threw-up-my-hands-michaels-story-of-overcoming-addiction-and-homelessness-through-faith-and-community>. Accessed 26 Jan. 2026.
- U.S. Department of Housing and Urban Development. "Homelessness and Employment." Cityscape, vol. 15, no. 3, 2013, <https://www.huduser.gov/periodicals/cityscpe/vol15num3/ch2.pdf>.

The Hidden Cost of Beauty: Shark-Derived Squalene and the Ethics of Cosmetic Transparency

By Lilian Thomas

Abstract

Despite their portrayal as fearsome predators, sharks are far more threatened by humans than they are a threat to us. Each year, an estimated 100 million sharks are killed, equivalent to three every second, with approximately 2.7 million of those deaths linked directly to squalene harvesting for the cosmetics industry. Squalene, a lipid found in deepwater shark livers, is valued for its hydrating properties in skincare but poses serious ethical and environmental concerns. This paper examines the ethical, environmental, and consumer-transparency issues surrounding shark-derived squalene in beauty products. It explores the history of squalene use, its biochemical properties in skincare, and the ecological consequences of declining shark populations. It then examines the transparency with which brands disclose to consumers the specific sources of squalene in their products. This study evaluates how transparent brands are in disclosing the sourcing of squalene using a dataset of 30 best-selling products from Sephora, a global beauty retailer that sells a wide range of skincare, makeup, haircare, and fragrance products. Findings reveal that 43% of products provide no information on whether their squalene is shark-derived or plant-based, limiting consumer awareness and informed choices. This study also reviews certifications such as “clean beauty,” “cruelty-free,” and “vegan,” highlighting inconsistencies and loopholes that enable greenwashing practices to thrive. This study emphasizes the urgent need for standardized certifications and stricter regulations.

Keywords

Shark, Squalene, Transparency, Conservation, Clean beauty, Cosmetic industry, Ethical and environmental concerns

Introduction

When a person pictures a shark, they likely imagine a ferocious killer preying on humans. This villainous image, highly perpetuated by the media, has led most people to accept the idea that sharks are bloodthirsty killers. In reality, in the relationship, humans are the real threat. Every year, humans kill an estimated 100 million sharks, or approximately 3 per second (1). This mass killing is done through a variety of means, including finning, unintentional bycatch, and for cosmetic use (2). This idea of the cosmetic industry may be a shock to many people. However, sharks are in high demand by large cosmetic businesses for a compound found in their livers, squalene. Every year, squalene harvesting is directly responsible for the death of an estimated 2.7 million sharks (2). One of the most concerning aspects of the squalene issue is the lack of consumer awareness. Without even knowing, many people, including those reading this article, have purchased a product without being thoroughly educated and made aware that a shark had to be killed for it. This two-syllable ingredient, squalene, seamlessly flows into the sea of black text on the back of products, likely never catching the eye of the average consumer.

Squalene is an organic compound primarily found in the livers of deep-sea sharks, where its low density aids in buoyancy regulation (3). Squalene makes up approximately 25% of a shark's body mass, and this buoyancy helps sharks work against gravity to avoid sinking or floating dangerously (3). Squalane is the hydrogenated, stable form of squalene, and it is the most commonly used synthetic form in products (4). The ingredient is highly valued in the beauty industry, particularly for its deeply hydrating and soothing properties, making it ideal for skincare and makeup. Squalane's ability to mimic the skin's natural oils makes it a popular emollient in various moisturizing products, particularly skincare products (4). As an emollient, it serves to hydrate the skin further. This moisturizing property has become increasingly in demand within the beauty industry in recent years, as concerns over air pollution and skin health have grown. Squalane works in skin care to strengthen and protect the skin's natural barrier, which helps fight bacteria and pollutants (5).

The use of squalane has had devastating consequences for shark populations, driving many species toward extinction. A majority of shark species are already on the verge of extinction due to human activities in general, and this demand for squalene has only further exacerbated that. The group most targeted for squalene harvesting is deepwater sharks, often found at depths of 400 meters or more. They are primarily targeted because their livers contain the largest amount of squalene (3). Over 60 shark species are hunted for their liver oil; among these, several are listed as "Vulnerable to Extinction" by the International Union for Conservation of Nature. Notable examples include the gulper shark, the Portuguese dogfish, the kitefin shark, and the basking shark.

The exploitation of sharks for squalene threatens not only individual species but entire marine ecosystems. As a keystone marine species, sharks play a vital role in maintaining ecological balance. Therefore, their population decline leads to devastation across the ocean's health.

In the last two decades, the use of squalene in the beauty industry has garnered greater attention as culture increasingly emphasizes the importance of "clean beauty." From this awareness, some activism has pushed brands to reconsider their sourcing practices. In the late 2000s, shark conservation activism, led by groups like Oceana, an environmental charity focused on raising awareness of human activities that actively harm the ocean and its inhabitants, led both Unilever and L'Oreal to publicly announce their shift away from shark-derived squalene. However, the obscurity of this product and the lack of education have made activism less prominent and successful.

This lack of education has made consumers unaware of what squalene is and, therefore, unable to understand why it is harmful. The obscurity has raised ethical concerns, prompting questions about what awareness brands owe consumers for this practice. This conversation around transparency has become more prominent in recent years as people have shifted toward seeking products that align with their values. This lack of understanding is one of the most significant factors driving this study, as it aims to raise awareness and examine brands' efforts to inform their consumers. This study aims to raise awareness and examine brand transparency

regarding squalene sourcing, as consumers increasingly seek alignment with ethical and sustainable brands.

Background

Squalene's discovery dates back to 1906, when Japanese researcher Mitsumaru Tsujimoto, an expert in oils and fats at the Tokyo Industrial Testing Station, isolated a high amount of unsaturated hydrogen from the liver of a deep-sea shark (6). This substance would be named “squalane” after the shark's scientific family, *Squalidae*. A decade later, in 1916, the extraction process was perfected.

When it was first discovered, it was used in factories as a machine lubricant (6). By the mid-1900s, researchers and industries began recognizing the compound's benefits for skincare. The compounds' hydrating properties made squalane especially valuable in Japan, where hydration-focused skincare was central to beauty routines.

The cosmetic industry did not gain much traction in the US until the early 1900s (7). As in the previous century, societal norms promoted a modest, natural appearance, and makeup was often stigmatized as something worn by actresses or prostitutes (7). Makeup saw a significant boom in the 1920s, with the term “makeup” coined by Max Factor in 1921 (7). With World War I forcing many women into the workforce, they became consumers, creating a market for products specifically targeted at them. In the 1940s and 1950s, cosmetics, including skincare products, experienced a second significant boom, largely driven by the aftermath of World War II. During this period, cosmetics became increasingly integrated into a person's daily routine (5). As it became more normalized and accessible, demand rapidly grew. This market growth led to a push for innovation as companies sought the most efficient ingredients. This demand for efficiency is what fueled the expansion of ingredients like squalane.

The mass expansion of the beauty industry brought challenges. Many of these issues stemmed from the still-relevant question: “What is in my products?” Early makeup products of the 1900s contained many hazardous chemicals unsuitable for human interaction. A significant event caused by this occurred in 1933 with the product Lash Lure, a permanent mascara, which left one woman dead and many others blind (8). This sparked mass hysteria among consumers who suddenly became interested in the safety of the cosmetic products they used. This culminated in the passing of the 1938 Federal Food, Drug, and Cosmetic Act. This act outlined that the Federal Drug Administration had the authority to approve ingredients and regulate cosmetics to ensure safety (8). The next considerable panic of ingredients occurred during the 1960s and 70s during the immense protest culture.

As a result of these protests, people became more aware of the environment and the dangers posed by products. People began to be concerned about the chemicals they might be exposed to and the unethical practices brands might use. This led to greater change from brands as they sought to align with these morals, pushing them towards slightly more sustainable practices. This is seen with the birth of clean beauty. The earliest examples of green and clean brands are seen with the founding of the Body Shop in 1976 and Aveda in 1978 (9). Both brands

represent a larger theme seen within the 70s, with brands taking on activism to gain consumers' favor. This activist movement is one of the first examples of consumers taking power from large brands and holding them accountable for their ethical and environmental actions. These ideas remained prevalent as the beauty industry expanded into the 21st century. The market globalized, and people kept their ideals about what they did and did not want within their products. As social media developed in the early 2000s, a large online culture emerged based on clean beauty. Influencers began to play a larger role in shaping the idea of what products were clean and which were not. This saw the largest development in consumer awareness, as detailed information about ingredients could be shared on a large, accessible scale.

With the movement toward clean beauty came the use of buzzwords and specific labels to signal brands' commitment to the cause. Some of the most popular labels seen on beauty products are vegan, cruelty-free, and clean. The issue with these labels is that they lack standardization, with many different definitions from various groups. The general definitions are as follows. The term vegan means the product contains no animal-derived ingredients (10). The term cruelty-free means the product does not test on animals, but it does not mean that the product is free from animal-derived ingredients (10). The term "clean" is least standardized, but it generally refers to products formulated without potentially harmful ingredients such as parabens, sulfates, phthalates, and synthetic fragrances (10).

While the rise of clean beauty has prompted greater awareness of ingredient sourcing, the lack of standardized definitions for terms like "clean," "green," and "natural" allows for ambiguity and misinterpretation. This ambiguity creates opportunities for greenwashing, the deceptive practice of marketing products as more environmentally friendly than they truly are. Brands use these terms loosely, without clear criteria or third-party verification, thereby misleading consumers seeking ethical and sustainable options. A study by Marta Gonçalves highlighted discrepancies between different brands' "clean" ingredient lists, demonstrating the inconsistency and potential for consumer confusion (10). Gonçalves analyzed how some brands labeled synthetic ingredients as "clean" while others excluded them. They noted that even the same ingredient could be accepted by one company and rejected by another based on internal, unpublished guidelines. The study also found that few brands disclosed their selection criteria transparently, making it difficult for consumers to understand what "clean" truly means across different labels. This lack of transparency erodes consumer trust and hinders their ability to make informed decisions aligned with their values.

With concern about more ethical ingredients spreading, brands are trying to find alternatives to ingredients like squalane. They began looking into eco-friendly sources of squalene from various plants. One such alternate source is olive-derived squalane, which has gained more prominence in the beauty industry in recent years. A difference between olive-derived squalane and shark-derived squalane is that olive-derived squalane concentration is significantly less, typically around 0.2% to 0.7%, compared to up to 80% in shark liver oil (11). Another top alternative source of squalene is sugarcane-derived squalene (12). It falls under the same issue as olive, but its saturation is significantly lower. The price of alternative sources is considerably higher than that of shark squalene due to lower saturation and a problematic extraction process. This makes sustainable squalane an extra expense, which makes it less

appealing to brands. Due to the opportunity cost of switching, many brands have chosen not to, sticking with shark-sourcing practices.

Literature review

Bloom conducted an extensive study into the use of squalene in 2012 and 2014. Their 2012 study found that the cosmetics sector mainly used animal squalene, even though alternative sources derived from sugar cane or olive oil were readily available. In 2014, the study tested 72 creams that listed squalene as an ingredient to measure any change. This study brought out a new idea for how brands get away with not marketing their sharks through blending, since “In more than 80% of the creams containing shark squalene, the squalene is a blend of plant and shark origin (the shark component often representing more than 50% of the blend)”. The study also found that many of the brands that used shark squalene claimed to be committed to environmentally friendly practices, and showed that consumers cannot always rely on brand messaging to make assumptions about brands' consistent practices.

A recent study by Nicholson (2024) examined the use of squalene in beauty products, finding significant discrepancies in labeling and squalene content. The study examined 157 samples from various brands (6). One part of the study took 88 samples from the top 20 global cosmetic parent companies. Using these samples, the study tested for shark squalene. The study found that 1 in 2 of the top 20 parent companies had at least one subsidiary brand that contained shark squalene. The study also examined the samples containing shark squalene and compared them to the ways the product was labeled, disturbingly finding “37% of samples found to contain shark squalene identified as vegan, either independently or through a recognized third-party certification, or made a definitive claim to use phytosqualene.”(6). This represents how, firstly, people are unable to trust the way brands are marketed towards them and, secondly, how labels like “vegan” may be inaccurate. The study also found that “89% of the samples containing sharks identify as cruelty-free. This statistic represents how brands can use buzzwords to manipulate customers. When brands claim they are cruelty-free, a consumer may imagine they are committed to treating animals safely, when in reality, this only means they do not test on animals directly. At the same time, they may still kill them for ingredients. The study found that 31% of products containing shark squalene were sourced from brands that identify as clean beauty. This represents a danger to consumers, as they may falsely trust these labels and believe they are making environmentally and morally conscious decisions through their purchases when, in reality, they are not. A notable drawback of the study is that there is no way to determine how sharks entered the product. That could have been done directly in mixing or, along the way, accidentally through manufacturing. This raises an ethical dilemma for the brands that may knowingly deceive consumers.

Environmentally, the use of shark squalene is hazardous. One of the main concerns about shark-derived squalene is its impact on shark populations. The shark population has rapidly declined in recent years, primarily due to human activities, including squalene extraction. As found by Pastoureau in 2021, “the global abundance of these predators declined by more than 70 percent due to such fishing.” Specifically, squalene derivation affects the deepwater shark

species, as deepwater sharks have a higher concentration of squalene, up to 80% of the total contents of oil within their liver (13). Deepwater sharks play a vital role in ocean ecosystems, serving the roles of recycling carbon by consuming dead animals on the ocean floor, maintaining ecosystem balance as apex predators, and protecting the health of ocean plants such as seagrass and reef plants (13). The slow extinction of these species will gravely affect ocean ecosystems, disrupting the balance and food chains of the creatures within them. Currently, half of the deep-sea shark species specifically targeted for the oil trade are threatened with extinction, and squalene's popularity is a significant contributor (13).

Environmental and ethical concerns

Another environmental concern is that as shark-derived squalene erodes shark populations, it removes one of the world's weapons against climate change. One of the main causes of climate change is carbon emissions, and in 2023, a new record was set at 419.3 parts per million (14). The ocean holds about 38,000 gigatons of carbon, which holds carbon against its release into the atmosphere. One of the largest holders is large sea creatures like sharks. According to the World Wildlife Fund, carbon is about 10%-15% of a shark's body. When land animals die, the carbon stored in their bodies is often released into the atmosphere during decomposition. However, this is not always the case with sharks. Many shark species, especially deep-sea ones, sink to the ocean floor when they die. In doing so, the carbon in their bodies can become part of the deep ocean ecosystem, where it may remain stored for long periods, a process known as carbon sequestration. When sharks are harvested for squalene, their bodies are removed from the ocean and processed, preventing this natural carbon storage and potentially releasing that carbon into the atmosphere instead.

Another way sharks combat carbon emissions is by protecting seagrass. Seagrass serves as a powerful carbon storage system that scientists believe can significantly mitigate climate change. The sediment of the seagrass absorbs the carbon and can hold it for decades or even centuries. According to Environmental Protection Agency biologist Peg Pelletier, "Seagrass meadows cover less than one percent of the ocean surface but contribute an estimated 10 percent of the annual carbon sequestration in the ocean," meaning it plays a pivotal role in the ocean's carbon mitigation (15). Sharks protect seagrass by preying on fish that eat it. Therefore, as shark populations decrease, fish populations increase, allowing the destruction of seagrass meadows.

There are many ethical concerns about the use of squalene. The first one is that its use contributes to the death of a living creature. When squalene is harvested, a shark must be killed, and there is no way to avoid this ethical dilemma when considering the topic of squalene. The use of shark-derived squalene forces a person to make the moral decision to prioritize the benefits of squalene over the life of an innocent animal.

Another ethical concern raised relates to corporate social responsibility. This relates to how much corporations owe consumers in terms of transparency in their relationships. It delves into the ethical question of how much exactly a company owes to consumers. It involves companies taking accountability for the impacts of their operations on employees, consumers,

communities, and the planet. When companies integrate corporate social responsibility into their business practices, it can help promote a positive image and build stronger relationships with consumers.

Materials and Methods

For my research, I focused on evaluating how transparent the source of squalene was for consumers. I used tools available to the average consumer to gauge the level of transparency regarding brands on this topic. I first visited the online retailer Sephora, one of the largest beauty retailers in the US. On the website, I used the search tool to look up “squalane” to find products that contained the ingredient. When I searched for this, there were 263 results, with the results falling in the various categories on the website. For the products, 200 were marked as in the skincare sector, 143 marked as in the clean at Sephora sector, 54 marked as in the Bath and Body sector, 57 marked as in the makeup sector, 22 marked as in the hair sector, and 2 marked as in the fragrance sector. I then sorted the results by bestselling and selected the top 30 products, skipping repeat brands. For each product, I measured how easy it was for consumers to determine which type of squalane was in the product to gauge how transparent the brands were. I first looked directly at the Sephora page for information on squalane sources in products and noted whether it was present. I then went directly to the product page on the brand's website to see if the information was available for the products that did not initially display it. I then again marked it down as present or absent. For brands where it was not present, I looked up “what type of squalane does (insert brand name) use” and saw if the brand had any statement on the ingredient. I then noted whether the remaining brands had it.

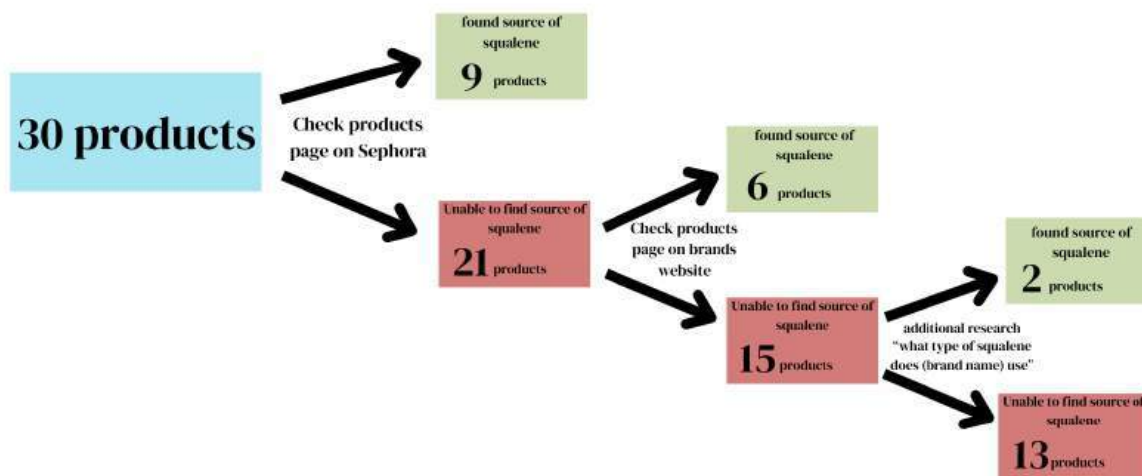


Figure 1. Visual aid for the method

ANALYSIS

In Figure 2, the y-axis lists the brands tested in the experiment. The four colors on the bar chart represent the steps needed to find the source of squalene. For brands with only green, the ingredient was available directly on Sephora's website, making it easy for consumers to find.

Brands that end in yellow indicate that the information was available on the brand's website, meaning it was still relatively accessible to consumers. Brands ending in orange indicate that the information was obtained through additional research, requiring consumers to conduct further research for an answer. For brands ending in red, the information was not accessible according to the study's measures. This means that for these products, the information would need to be obtained through additional research measures, or it may not be accessible. The left graph in Figure 2 shows the listed brands and the number of levels required to find the results. The right graph in Figure 2 ranks products by their position on Sephora's best-selling list.

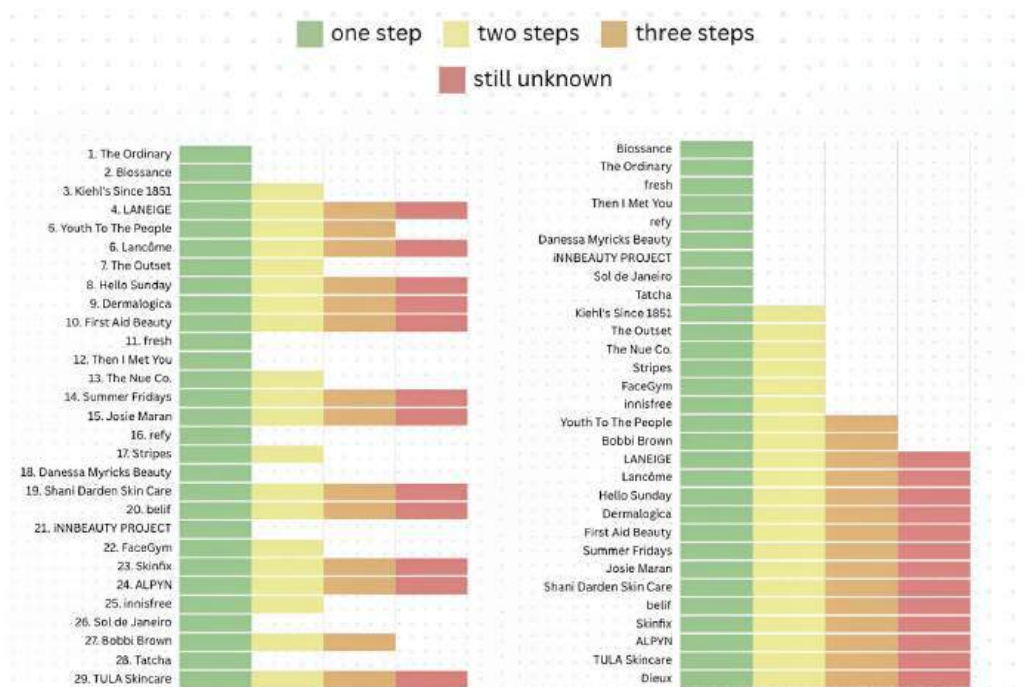


Figure 2. Top 30 brands that appeared when Squalane is searched in Sephora in November 2024

The results of the data analysis show an apparent lack of transparency over squalane sourcing. While brands make it clear in the ingredient section that Squalane is used, there is little to no information about the type of Squalane used. This issue is prevalent in both global beauty giants and smaller businesses, underscoring that transparency is a challenge across the board.

Looking at the data, 30% of the products on the Sephora page indicated sourcing for their squalane, while 70% did not. Of the 70% that required additional information, 29% were available on the brand's website for the product, and 71% required further research. Of the 71%, 13% were identified through further research into brands' public statements, and the rest remained unknown. Of the 30 tested products, 43% lacked information on the type of squalane used. This raises alarming issues, as it means only a little more than half the products were actually accessible to consumers, so they could know what kind of squalene was used. This is the

difference for a consumer in knowing whether they are making purchasing decisions that align with their morals.

Two of the unavailable sources were among the top ten. These are the top ten brands supposedly most popular and reliable, but two of them lack transparency with consumers. This is concerning, as these are the brands consumers are most likely to purchase, and they should all uphold transparency standards.

An interesting note in the data analysis was that the top two brands recommended by Sephora when the keyword “squalane” was searched were “The Ordinary” and “Bioessence.” These brands are pivotal members of the authentic, clean beauty movement and large advocates of clean squalane. When Squalane was searched, the top two products recommended from each brand were The Ordinary’s “100% Plant-Derived Squalane” and Bioessence’s “100% Sugarcane Squalane Oil”. The titles of these products provide clear guidance to consumers, serving as a positive example of how brands can be transparent. The brand Bioessence also has a page on its direct website that provides information on squalane and denounces its unethical sourcing practices through a partnership with the charity Oceana. These brands serve as an example of what genuine commitment to social movement in brands can be seen, rather than a superficial one. Their high recommendation also offers hope and proof that consumers prefer transparent products and brands committed to social issues.

A key aspect of the research is that every product demonstrating transparency was found to utilize an ethical source. No brands explicitly claimed to use shark-derived squalene within their advertising. As shown in Figure 3, of the 17 transparent brands, three used olive-derived squalane, five used sugarcane-derived squalene, six used plant-derived squalane, one used biomimetic-derived squalene, and two listed their source as just “vegan.” Even here, with the only term given as 'vegan,' it shows that there is still ambiguity about the type of squalene used. While there is no way to prove it, there is a higher chance that brands that do not disclose the kind of squalene used directly are using shark-derived squalane. This is hinted at as it is something they would be more likely to hide, as brands would not feel shame in sharing if it were a less controversial practice. This may also suggest that brands are not sufficiently transparent with consumers and do not consider sourcing information vital to include. This shows a lack of care for consumers on the part of brands, as they are completely unclear about their sourcing.

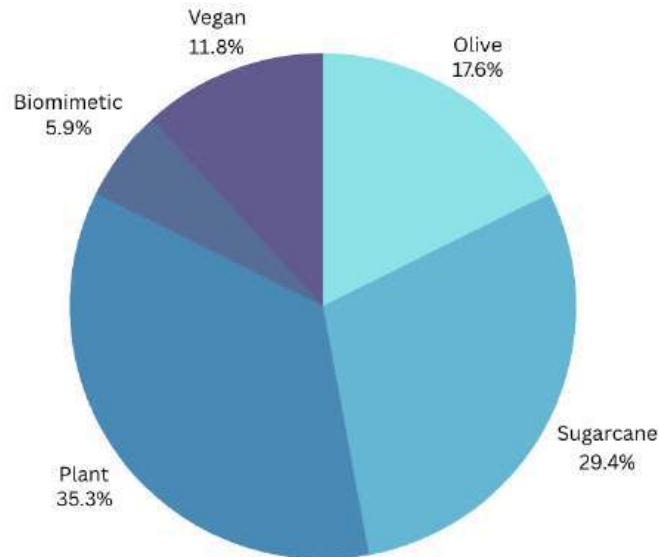


Figure 3. Type of Squalene source used in products, where information was available as of November 2024, according to Sephora

In this data, it must be noted that just because the information on the squalene source was unavailable, it does not automatically mean the products contain shark squalane. Instead, it highlights uncertainty, emphasizing that brands should clearly distinguish between their clean and unclean practices for consumers. A limitation in data collection is that assumptions about the product's contents can be made only from marketing information rather than from accurate laboratory testing. This means that there must be trust given to brands that claim to use a particular source, as there is no way in the study to test their truth. This mimics a consumer's resources when purchasing a product, and the same limitations apply when determining what is in a product.

This raises the significant issue of corporate social responsibility, prompting a question about the extent to which brands owe consumers ethicality and transparency. In this situation, it raises the question of whether brands owe consumers this information and how brands can be held accountable for the lack of clarity. The grave danger of consumers not being aware of precisely what is in their products is that it undermines their ability to make informed, personal ethical decisions when purchasing. Without understanding the practice behind a product, a consumer's ability to follow their moral code is limited when making a purchase.

DISCUSSION

Based on this research's findings, further regulation is necessary in the beauty industry. As evidenced by the study, many brands make little to no effort to provide proper clarity to consumers about what is in their products, leaving consumers at a dangerous disadvantage. The lack of clarity extends beyond environmental impact and can be hazardous to those with allergies

or other factors that necessitate avoiding specific ingredients. There must be more regulation to ensure brands make transparent information available. Also, based on findings from previous studies, it is clear that there is a lack of regulation regarding the marketing of these products through buzzwords. The Federal Drug Administration should standardize these terms to ensure consumers understand their true meanings and that brands are held accountable for upholding them. Brands must also comply with regulations regarding how their activities may negatively impact the environment, and these protocols prioritize protecting the natural world.

For brands, transparency is essential as it fosters a stronger connection with consumers. It would be beneficial for brands to be transparent about their ingredients, ensuring that consumers know what they are using. A recommendation would be to add pages on the website that explain the different ingredients and the types the brand uses. This could specifically follow the examples set by websites such as Bobbi Brown's cosmetic glossary, which clearly explain their key ingredients and sources. Another example is the Bioessence website, which directly explains what squalene is and outlines their process. Through pages like this, brands can provide greater clarity about their products, thereby fostering transparency and attracting customers.

For consumers, it is recommended that they pay attention to which brands they support and investigate their practices. They should also hold businesses accountable and remember that advocacy starts with the customer. They should not blindly trust labels; instead, they should investigate the ingredients in their products.

CONCLUSION

The issue of Squalene's use in the beauty industry highlights an urgent need for attention and education about sometimes unethical practices. The issue with Squalene is that most consumers are unaware of its use and its role in environmentally irresponsible actions. This allows brands to harm their consumers and fail to respect corporate social responsibility. Despite the availability of sustainable alternatives, the widespread use of shark-derived squalene in cosmetic products contributes to the ongoing decline of shark populations. It threatens the delicate balance of marine ecosystems and contributes to the ongoing death of living creatures. This case shows the need for transparency and accountability for large brands that hold all the power in the relationship. It is found that few brands provide a clear answer about the source of their squalene. This issue, however, is likely applicable to other ingredients used across the beauty industry that consumers should be aware of. The issue underscores why transparency is vital and why it is time for brands to give consumers what they deserve.

Works Cited

1. Hopkins, Rachel, and KerriLynn Miller. "Alarming Decline in Shark Populations Points to Need for Stricter Conservation Measures." Pew, 9 Mar. 2021, www.pewtrusts.org/en/research-and-analysis/articles/2021/03/09/alarming-decline-in-shark-populations-points-to-need-for-stricter-conservation-measures.
2. BLOOM Association. *Shark in Our Beauty Creams: An Investigation into the Market of Deep-Sea Shark Liver Oil*. 2012, www.bloomassociation.org/en/wp-content/uploads/2018/04/squalane-bloom-english-1.pdf.
3. "Biochemical and Physiological Adaptations to Depth in Deep-Sea Sharks." web.uri.edu/wetherbee/biochemical-and-physiological-adaptations-to-depth-in-deep-sea-sharks/.
4. Clinic, Cleveland. "What Is Squalane?" Cleveland Clinic, 23 Dec. 2022, health.clevelandclinic.org/squalane.
5. Huang, Ying-Ying, et al. "Explicating the Multifunctional Roles of Tocotrienol and Squalene in Skin Health: A Review." *Skin Health and Disease*, vol. 3, no. 3, 2023, e448. Wiley Online Library, <https://onlinelibrary.wiley.com/doi/full/10.1002/ski2.448>.
6. Nicholson, Sarah. *Sharkless Makeup > Sharkless Oceans*. June 2024, <https://doi.org/10.32920/26053096.v1>.
7. Smithsonian Institution. "Cosmetics and Personal Care Products in the Medicine and Science Collections." Smithsonian Institution, 2016, www.si.edu/spotlight/health-hygiene-and-beauty/make-up.
8. Commissioner, Office of the. "The 1938 Food, Drug, and Cosmetic Act." FDA, vol. 2, Apr. 2019, www.fda.gov/about-fda/histories-product-regulation/1938-food-drug-and-cosmetic-act.
9. "Our Story - Heritage." Aveda, www.aveda.com/horst-rechelbacher-aveda-history. perfect point. "What Does Clean Beauty Mean? An In-Depth Analysis of Key Terminology in the Beauty Industry - Avenue Five Institute." Avenue Five Institute, Nov. 2021, www.avenuefive.edu/what-does-clean-beauty-mean/.
10. Anderson, Lisa. "As Growing Demand for Squalene Leads to More Shark Deaths, Olives Offer a Solution." *Olive Oil Times*, 19 July 2021, www.oliveoiltimes.com/world/as-growing-demand-for-squalene-leads-to-more-shark-deaths-olives-offer-a-solution/96529.
11. "Deepwater Sharks Are Threatened by Demand for Liver Oil - Saving Seafood." *Saving Seafood*, 9 Mar. 2024, www.savingseafood.org/news/conservation-environment/deepwater-sharks-are-threatened-by-demand-for-liver-oil/.
12. Pastoureaux, Nathan, et al. "Half a Century of Global Decline in Oceanic Sharks and Rays." *Nature*, vol. 589, no. 7843, Jan. 2021, pp. 567–71, <https://doi.org/10.1038/s41586-020-03173-9>.

13. National Oceanic and Atmospheric Administration. "Greenhouse Gases Continue to Skyrocket, Hitting Record-High, NOAA Says." Axios, 8 Apr. 2024, www.axios.com/2024/04/08/greenhouse-gases-record-noaa.
14. US EPA, ORD. "EPA Scientists Study the Carbon-Storing Power of Seagrass to Fight Climate Change." Wwww.epa.gov, 1 Feb. 2022, www.epa.gov/sciencematters/epa-scientists-study-carbon-storing-power-seagrass-fight-climate-change.

Aesculin: Natural Functions, Unique abilities, Current and Future Medical Possibilities.

Viacheslav Detkov

Abstract

Aesculin is a coumarin glucoside which is naturally found in *Aesculus hippocastanum* and *Aesculus chinensis* has been a part of traditional European herbal medicine and was previously regarded as a quality-control marker and not an active medical compound. Recent studies have highlighted the significance of Aesculin, giving insights into its antioxidant, anti-inflammatory, antiviral, neuroprotective and antitumor effects.

Literature data about biochemical, genomic, and clinical perspectives, suggest that Aesculin functions as a multifunctional protective compound within the plants and is a bioactive molecule used for pharmaceutical and medical applications.

The genomic identification of coumarin-glycoside biosynthetic pathways in *Aesculus* species supports future possibilities for biotechnological synthesis and a great potential in further pharmaceutical development. Motivating further research into its molecular mechanisms, optimized formulations and possible medical applications.

This review discusses current findings about the compound, its natural occurrence, chemical characteristics, pharmacological history, and future medical potential of Aesculin.

Introduction

Aesculin (Fig. 1), also known as Esculin, exhibits toxicity in raw or unrefined forms but therapeutic activity at controlled doses. It is found in the European Horse Chestnut trees (*Aesculus hippocastanum* L.) (Stanić et al., 1999) and Chinese Horse Chestnut trees (Dudek-Makuch & Matlawska, 2013). Aesculin has emerged as one of its pharmacologically significant constituents and is chemically it defined as 6-O-β-D-glucopyranosyl Aesculetin, combining the phenolic antioxidant core of Aesculetin with a glucose moiety enhancing the solubility, bioavailability and biochemical versatility of the compound (Stanić et al., 1999; Shahidi & Ambigaipalan, 2015).

Biological and therapeutic effects of Aesculin are no longer a topic of research with Aesculin mostly being used as a chemical indicator confirming that a horse chestnut extract is genuine, properly prepared and consistent chemically. Aesculin is recognized as a bioactive molecule with therapeutic potential, expanding beyond the traditional uses of the compound. Modern pharmacological research has shown Aesculin's ability to modulate oxidative stress, inflammation, viral infection and neuronal injury, mechanisms crucial to underpin various diseases, including venous insufficiency and neurodegeneration (Cai & Cai, 2023; Ju et al., 2024; Thangarajan et al., 2023). Aesculin has been found incorporated into various medical formulations, such as standard horse chestnut extracts for circulatory disorders and dry eye diseases (Underland et al., 2012; Wang et al., 2025). These uses of Aesculin are clearly

demonstrating the clinical potential of the natural compound, however well established safety precautions are needed.

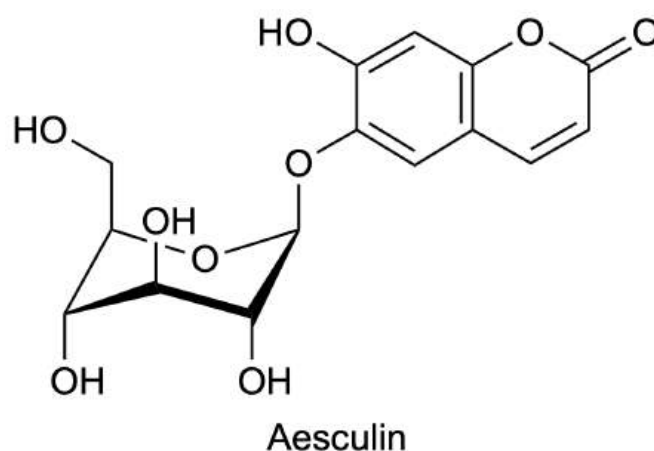


Figure 1. Chemical structure of Aesculin.

Multiple genomic and biochemical studies have been performed on the *Aesculus* species, specifically looking at Aesculin synthesis through coumarin-glycoside biosynthetic gene clustering. This pathway evolved through whole-genome duplication and enzymatic specialization (Sun et al., 2023), providing a molecular foundation for the biotechnological production and the drug development with Aesculin and its derivatives present at scale.

Though Aesculin has shown promising developments in recent years, the compound still is fairly understudied compared to its full pharmacological potential. Data on pharmacokinetics, cellular targets, and long-term clinical effects remain limited. (EMA, 2011). There are cases of toxic poisoning from consumption of the raw *Aesculus* seeds (*Aesculus hippocastanum* L.) which remind researchers about the dosage control importance (Yi & Lee, 2021). Standardized extracts with formulated preparations show that the toxicity of Aesculin can be avoided in pharmaceutical ranges.

Aesculin is both a traditional healing compound and a bioactive agent. This promising compound provides an important bridging between natural herbal products and a modern medicine medalist. The purpose of this review is to raise awareness of Aesculin's importance in the biomedical field, as well as to discuss and highlight its future potential as a safe, non-toxic pharmaceutical for clinical and medical uses.

Main text

1. Natural Occurrence, Ecological Role and Properties of Aesculin

Aesculin is found naturally occurring in the species of *Aesculus hippocastanum* L. and *Aesculus chinensis* with the highest concentrations found in the bark, flowers and seeds. As

well as Aesculin, Aesculetin and Fraxin are also found in these species (Stanić & Jurić, 1999.; Dudek-Makuch & Mat, 2011). According to the European Medicines Agency, *Aesculus hippocastanum L.* bark typically contains around 4.4-7.7% of dried Aesculin material. This relatively high abundance suggests that Aesculin plays an important physiological role in the plant rather than being a metabolic by-product.

A known property of Aesculin is its ability to emit a light blue glow under ultraviolet (UV) light, due to the presence of coumarin residue (Shahidi & Ambigaipalan, 2015). This was demonstrated through a simple experiment conducted for this literature review. A freshly cut Horse Chestnut (*Aesculus hippocastanum L.*) branch was placed into a beaker containing 220 ml of water, where the branch was gently stirred to allow water-soluble compounds to diffuse into the solution. When the beaker with the resulting solution was exposed to UV light, it emitted a visible blue fluorescence, indicating the presence of a UV-active compound, which was likely Aesculin (Fig. 2 and Fig. 3).



Figure 2. A beaker containing 220 ml of water, exposed to UV light lamp ($\lambda = 254 \text{ nm}$). No fluorescence observed due to absence to UV-active components in solution.



Figure 3. A beaker containing 220 ml of aqueous solution after Horse Chestnut Tree branch stirred, exposed to a UV light lamp ($\lambda = 254 \text{ nm}$), showing blue fluorescence indicating the presence of Aesculin.

Genomic analysis of *Aesculus chinensis* has identified gene clusters responsible for the biosynthesis of coumarin-glycoside (Sun et al. 2023). These genes are also a part of the evolutionary stable secondary-metabolic structure found in the *Aesculus* species. Most likely this developed in order to produce phenolic compounds, protecting the trees against stress from oxidation and pathogens. Previous research also suggests additional evidence of Aesculin being a part of a defensive mechanism of *Aesculus* tree species (Dudek-Makuch and Mat 2011).

The antioxidant potential of Aesculin was described in several studies (Ju et al. 2024, Cai, 2023), reinforcing the uses of it in the ecological systems. In both of these studies, it was observed that coumarines neutralise reactive oxygen species in biological models. In the plant, these antioxidant reactions contribute to cellular protection from UV radiation and oxidative damage and stress caused by photosynthesis. Furthermore, Aesculin and other coumarins can go into enzymatic oxidation and polymerization (Muñiz-Mouro et al. 2017) using laccase-mediated reactions in *Aesculus* extracts. These enzymatic changes may enhance the ability of plants to form polymeric barriers in response to environmental damage.

In conclusion, these observations and findings indicate the multifunctional purpose of Aesculin in the trees, functions essential for survival and resilience of *Aesculus* species.

2. Pharmacological History and Medical Use of Aesculin

The use of Aesculin in medicine has been rooted in the long-standing therapeutic use of *Aesculus hippocastanum*. The bark and seeds of these trees have been used in European herbal medicine for centuries to relieve venous insufficiency, inflammation and swelling (EMA, 2011; Underland, Sæterdal, & Nilsen, 2012). Early medicinal preparations were mostly standardized for Aescin, but later it was found that Aesculin and its aglycone, Aesculetin, were also major contributors to the pharmacological profile of horse chestnut extracts (Stanić & Jurić, 1999; Dudek-Makuch & Mat, 2013).

First time the compound has been isolated, was partially an effort to identify the active coumarines present in *Aesculus* species. Later on Aesculin became a reference and a quality-control marker for horse chestnut extracts, due to abundance of the compound and how easy it was to detect it by HPLC and UV spectrophotometry (EMA, 2011; Edberg, Pittman, & Singer, 1977).

Throughout the 20th century, Aesculin was of high medical value, and was further recognised to have antioxidant and anti-inflammatory effects in the plant's vascular and protective functions (Cai & Cai, 2023; Ju et al., 2024). Later on, it was emphasized that coumarins derivatives such as Aesculin have a potent phenolic antioxidant activity, which explained the use of it for circulation and tissue repair (Shahidi and Ambigaipalan, 2015). Aesculin has been found to also have neuroprotective effects in lead-induced toxicity (Thangarajan et al. 2023) while antiviral efficacy of Aesculin was also established in aquatic disease models (Shan et al. 2022 and Xue et al. 2024). These findings expanded the abilities of Aesculin as a pharmaceutical, establishing it as a multifunctional therapeutic compound.

The clinical applications of Aesculin have since validated the safety and efficacy of it. After conducting a randomized controlled trial, Aesculin and digitalis-glycoside-based eye drops improved tear stability and ocular surface health in patients with dry eye disease (Wang et al. 2025). The continued use of Aesculin in EMA-approved herbal monographs (EMA, 2011.) and according to evidence based reviews confirmed its integration into regulated modern medicine (Underland et al. 2012).

Ever since its discovery as a fluorescent coumarin glycoside to its uses in pharmacological applications, this progress reflects the transition of natural compounds like Aesculin from traditional remedies to scientifically validated medications.

3. Future Medical Potential of Aesculin

Recent research studies have shown that Aesculin shows a great potential for future broad medical application far beyond the previously established vascular and anti-inflammatory uses. Recent neuroprotective results showed that Aesculin reduced lead-induced neurotoxicity in the hippocampus and cortex of mice. This suggests that the compound is able to be a candidate for preventing or managing the neurodegenerative disorders linked to oxidative stress (Thangarajan et al. 2023). In addition, it was demonstrated that a related coumarin previously mentioned – Aesculetin – provided neuroprotection in Huntington's disease models, showing that Aesculetin and its glycosylated analogue –

Aesculin – have comparable benefits of better solubility and stability (Pruccoli et al. 2021). The antiviral activity of the Aesculin against white spots syndrome and against largemouth bass ranavirus highlights the wide spectrum potential of the compound as an antiviral agent, capable of replicating and maximizing the host's immune response to triggers (Shan et al. 2022 & Xue et al. 2024)

Anti-tumor effects of Aesculin were also proved in *in vitro* studies (Liu et al. 2024 & Hossain et al. 2024). These findings suggested Aesculin as a promising multi-target compound with possible potential in oncology, neurology, virology, and ocular medicine.

Advances in plant genomics have further broadened future possibilities. The identification of coumarin-glycoside biosynthetic genes in *Aesculus* species opens opportunities for biotechnological production or genetic engineering of Aesculin analogues with enhanced pharmacokinetic profiles (Sun et al. 2023). Future research and testing should focus on elucidating Aesculin's molecular targets, optimizing delivery systems, and conducting more clinical trials in order to verify its therapeutic potential. All together the evidence from molecular case studies, animal models and clinical observations support the evolution of Aesculin from a traditional plant used for more than 2000 years to a versatile bioactive compound with strong promises for future development in the pharmaceutical field.

Conclusions and Outlook

Aesculin is a strong example of traditional herbal medicine turning into a modern medical prospect. Originally identified as natural coumarin glycoside in *Aesculus hippocastanum* and *Aesculus chinensis*, it was long regarded mainly as a quality-control marker for horse chestnut tree extracts. However, after years of medical research it was shown to have great potential in a broad range of biological activities, including antioxidant, anti-inflammatory, neuroprotective, antiviral and antitumor effects (Cai & Cai, 2023; Ju et al., 2024; Thangarajan et al., 2023; Shan et al., 2022; Xue et al., 2024; Liu et al., 2024; Hossain et al., 2024).

Due to its distinctive photochemical reactivity, genomic basis of biosynthesis, and proven pharmacological efficacy, Aesculin demonstrates as a versatile bioactive molecule with therapeutic value across various systems. Traditional uses of *Aesculus extracts* are still being used in vascular diseases treatment, modern medical studies have greatly extended the possibilities of Aesculin. Nevertheless there are still gaps in the understanding of pharmacokinetics, optimal dosage, and long-term safety of Aesculin as a medical agent.

To fully understand the medical potential of Aesculin, future research is needed, including molecular studies and biotechnological approaches. The growing amount of evidence underscores Aesculin as a compound that bridges traditional herbal medicine with modern pharmacology reminding us that some very promising medical innovations continue to originate from nature.

Bibliography

1. Cai, T., & Cai, B. (2023). *Pharmacological activities of esculin and esculetin: A review. Medicine*, 102(40), e35306. <https://doi.org/10.1097/MD.00000000000035306>
2. Dudek-Makuch, M., & Mat, I. (2012). *Coumarins in horse chestnut flowers: Isolation and quantification by UPLC method*.
3. Edberg, S. C., Pittman, S., & Singer, J. M. (1977). *Esculin hydrolysis by Enterobacteriaceae. Journal of Clinical Microbiology*, 6(2), 111–116. <https://doi.org/10.1128/jcm.6.2.111-116.1977>
4. European Medicines Agency (EMA). (2023). *Assessment report on Aesculus hippocastanum L., cortex*. Committee on Herbal Medicinal Products (HMPC). https://www.ema.europa.eu/en/documents/herbal-report/draft-assessment-report-aesculus-hippocastanum-l-cortex_en.pdf
5. Hossain, M. R., Shova, F. T. Z., Akter, M., et al. (2024). *Esculetin unveiled: Decoding its anti-tumor potential through molecular mechanisms—A comprehensive review. Cancer Reports*, 7(1), e1948. <https://doi.org/10.1002/cnr2.1948>
6. Ju, S., Tan, Y., Wang, Q., et al. (2024). *Antioxidant and anti-inflammatory effects of esculin and esculetin (review). Experimental and Therapeutic Medicine*, 27(6), 248. <https://doi.org/10.3892/etm.2024.12536>
7. Liu, M., Sheng, Y., Guo, F., et al. (2024). *Therapeutic potential of esculetin in various cancer types (review). Oncology Letters*, 28(1), 305. <https://doi.org/10.3892/ol.2024.14438>
8. Muñiz-Mouro, A., Oliveira, I. M., Gullón, B., et al. (2017). *Comprehensive investigation of the enzymatic oligomerization of esculin by laccase in ethanol: Water mixtures. RSC Advances*, 7(61), 38424–38433. <https://doi.org/10.1039/C7RA06972C>
9. Pruccoli, L., Breda, C., Teti, G., Falconi, M., Giorgini, F., & Tarozzi, A. (2021). *Esculetin provides neuroprotection against mutant huntingtin-induced toxicity in Huntington's disease models. Pharmaceuticals*, 14(10), 1044. <https://doi.org/10.3390/ph14101044>
10. Shahidi, F., & Ambigaipalan, P. (2015). *Phenolics and polyphenolics in foods, beverages and spices: Antioxidant activity and health effects – A review. Journal of Functional Foods*, 18, 820–897. <https://doi.org/10.1016/j.jff.2015.06.018>
11. Shan, L. P., Zhang, X., Hu, Y., Liu, L., & Chen, J. (2022). *Antiviral activity of esculin against white spot syndrome virus: A new starting point for prevention and control of white spot disease outbreaks in shrimp seedling culture. Journal of Fish Diseases*, 45(1), 59–68. <https://doi.org/10.1111/jfd.13533>
12. Stanić, G., & Jurić, B. (n.d.). *HPLC analysis of esculin and fraxin in horse-chestnut bark (Aesculus hippocastanum L.)*.
13. Sun, W., Yin, Q., Wan, H., et al. (2023). *Characterization of the horse chestnut genome reveals the evolution of aescin and aesculin biosynthesis. Nature Communications*, 14(1), 6470. <https://doi.org/10.1038/s41467-023-42253-y>
14. Thangarajan, S., Murugan, A., Rajavel, S., & Perumal, S. I. (2023). *The antioxidative effect of esculin on lead acetate-induced neurotoxicity in the hippocampus and cortex*

- of C57BL/6 mice. *International Journal of Pharmaceutical Sciences and Drug Research*, 15(4), 399–406. <https://doi.org/10.25004/IJPSDR.2023.150401>
15. Underland, V., Sæterdal, I., & Nilsen, E. S. (2012). *Cochrane summary of findings: Horse chestnut seed extract for chronic venous insufficiency*. *Global Advances in Health and Medicine*, 1(1), 122–123. <https://doi.org/10.7453/gahmj.2012.1.1.018>
 16. Wang, Y., Yuan, J., Wen, Y., Li, S., Tian, L., & Jie, Y. (2025). *Clinical observation of esculin and digitalis-glycosides eye drops with 0.3% sodium hyaluronate eye drops for dry eye disease: A randomized controlled trial*. *Scientific Reports*, 15(1), 5747. <https://doi.org/10.1038/s41598-025-90074-4>
 17. Xue, M., Li, J., Meng, Y., et al. (2024). *Antiviral effects of esculin on largemouth bass ranavirus in vivo and in vitro*. *Aquaculture Reports*, 34, 101882. <https://doi.org/10.1016/j.aqrep.2023.101882>
 18. Yi, H. Y., & Lee, J. Y. (2021). *Poisoning due to consumption of horse chestnut seed*. *Clinical and Experimental Emergency Medicine*, 8(4), 333–335. <https://doi.org/10.15441/ceem.20.004>

How Cook County's Complex Bureaucracy Worsens Inequity and Compromises Public Trust By Kayla Moore

When homelessness, 50 manual school closures, and housing insecurity becomes more prevalent in a county, the citizens often blame and resent the local government. However, when parts of the bureaucratic administration responsible begin to reform its system in order to help citizens gain money back, keep their homes, and make them aware of the intricacies to the bureaucracy (a system of government where most decisions are made by state officials not elected representatives according to the Dictionary), public trust could be regained. Beyond reasons such as inflation and overpopulation, the complexity of the bureaucracy has also caused property taxes to increase and public schools to be at risk. Cook County's public administration and bureaucracy can best be described as following Weber's bureaucratic rational/legal model or the Weber Theory: Authority in governments comes from legal rules, regulations, and those in professional positions, as opposed to tradition or loyalty to a powerful figurehead¹ – such as the military, large businesses, or certain bureaucracies. However, with a plethora of intricate aspects to Cook County's bureaucracy and the thousands of Cook County official jobs, the complexity of its bureaucracy can hinder positive economic and social development of its citizens.

In context, Cook County is an urban county in the northeast of Illinois that has a population of 5.3 million people, more than 800 local governmental units and is the 19th largest government in the United States (2015 census statistics).² Cook County is governed by a 17-member Board of Commissioners elected from single-member districts. The Commissioners and a County Board President (president of the governing legislative body in Cook County) are elected for four-year terms by Cook County's citizens³ according to the Cook County government. It contains 134 municipalities, with its most well-known being the City of Chicago. "As mandated by State law, the County government has principal responsibility for the protection of persons and property, the provision for public health services and the maintenance of County highways."⁴ Navigating where to allocate the County's money to and how to create equity seem to be arduous tasks with many obstacles: A solution to generate equity and ensure

¹ Cook County Treasurer's Office. "Duties and Responsibilities." n.d.

² Cook County Government. "About Cook County." n.d.

³ Ibid.

⁴ Ibid.

future reform must be created to generate public trust among the citizens and foster more frequent civic engagement in political elections through increased turnout.

How does the bureaucratic complexity of Cook County's government influence public trust in local governance?

Cook County's complex bureaucracy mostly hinders public trust in local governance. While the Assessor's Homeowner Relief Fund and outreach efforts by one county official aim to address inequity, the three-year reassessment cycle and uncoordinated network of smaller governments and departments sustain property-tax inequities, lower appeal (an application to higher authority for a decision to be reversed according to the Dictionary) rates in minority neighborhoods, and the manual shutdown of many Chicago public schools. This instability fuels mistrust and broadens voter-turnout gaps between wealthier neighborhoods protected by the bureaucracy and the minority, substandard neighborhoods it harms.

Firstly, the Cook County Assessor is the elected official responsible for establishing fair and precise property assessments for the county's 1.8 million parcels (identifiable portions of land) for tax purposes. The Cook County Assessor's office uses a three-year reassessment cycle to measure property changes for permit applications, property division, demolition, and other applications. This year, the south suburbs were assessed which showed large valuation jumps for parcels (specific, defined pieces of property recognized as distinct entities for legal ownership, land use, and taxation⁵ according to investopedia) and significant increases in yearly tax bills. The Treasurer showed median south-suburban residential tax bills rose by 18-20% in the last two years, and county relief programs were now aimed at households whose bills rose by 50% or higher in a single year according to the Cook County Assessor's Office.⁶ As the three-year reassessment cycle neglects certain suburbs like the south for the other two years unassessed, rapid reassessment can cause tax pressure on long-time homeowners in low-income communities because they are unaware of the tax increases and are not provided enough relief programs⁷ or outreach help according to WBEZ Chicago. Therefore, the three-year reassessment structure used by the Cook County Assessor's office is ineffective and provides instability to low income homeowners in the south suburbs especially.

⁵ Investopedia. "Property Tax." n.d.

⁶ Cook County Assessor's Office. "About Cook County Assessor's Office." n.d.

⁷ WBEZ Chicago. Reporting on Cook County Reassessments. n.d.

Secondly, the Cook County Treasurer oversees the property-tax collection and distribution system. Duties include printing and mailing bills; collecting \$16.1 billion annually from more than 1.8 million parcels; and distributing funds to approximately 2,200 local government agencies (Cook County Treasurer's Office). The treasurer also allows appeals (a request for a higher authority such as a court or governmental body to review and change a decision made by a lower authority) filed by citizens in order to measure satisfaction on property taxation in Cook County according to the official appeal process.⁸ Yet, the treasurer is not responsible for allocating property taxes. Cook County analysis reveals significant varied appeal rates from neighborhoods based on differing financial incomes and race. For example, there was a 46% appeal rate in the highest income neighborhoods and an 11% appeal rate in the lowest-income neighborhoods. In predominantly white neighborhoods, appeal rates were 35.5%, 10.5% in predominantly Black neighborhoods, and 14% in mostly Latino neighborhoods according to the Cook County Treasurer's office.⁹ The disproportionality of appeal rates highlights the underlying lack of awareness on the appeal process based on race and economic income and the fact that appeals made by high-income white people are more likely to secure assessment reductions than those made by Black or Latino people because they are more well-versed in the intricate tax system. Furthermore, a study conducted by the Cook County Treasurer showed how the tax appeals system tends to favor appeals made by the same white and high-income people and favors their neighborhoods; whereas, appeals made by minority low-income neighborhoods can cause these neighborhoods financial damage. In a Census tract in Gage Park, a majority Latino and low-income community location on the Southwest Side saw their tax bills rise by almost 23% after only 5.2% of homeowners filed for appeals. On the other hand, the Census tract in North Center, a predominantly white and wealthy neighborhood, saw a tax bill rise of less than 15% when 60% of the homeowners appealed their tax assessments according to Cook County's Treasurer's Office.¹⁰ Therefore, appeals made by lower income minority populations in Cook County are less likely to be successful in reducing tax bills than

⁸ Lake County, Illinois. "Appeal Process." n.d.

⁹ Cook County Treasurer's Office. "The \$2 Billion Wealth Transfer: How Cook County's Property Tax Appeals System Rewards Businesses at the Expense of Homeowners." *Cook County Treasurer's Office*, May 5, 2025.

¹⁰ Cook County Treasurer's Office. "The \$2 Billion Wealth Transfer: How Cook County's Property Tax Appeals System Rewards Businesses at the Expense of Homeowners." *Cook County Treasurer's Office*, May 5, 2025.

those made by wealthy white neighborhoods and actually increase their tax bills and worsen inequity.

However, Cook County uses the Assessor's Homeowner Relief Fund in its bureaucracy. It is used in conjunction with the Cook County Treasurer and multiple community organizations to increase awareness and improve inequity in Cook County according to the Cook County Homeowner Relief Fund.¹¹ This fund functions in multiple ways: to assist homeowners, mainly in the west and south suburbs, struggling financially due to increased property taxes; the fund gives outreach support and grants to aid residents' applications for exemptions, appeals, and accessing refund programs; the fund supports low-income, long-time residents, and senior citizens who are close to losing their homes because of financial hardship; and the Assessor's Homeowner Relief Fund reduces financial strain by showing how the government can respond to residents affected by inequity according to the Chicago Sun-Times.¹²

Furthermore, Cook County Treasurer Maria Pappas created the "Black Houses Matter" initiative to find, recognize, and refund property tax overpayments in Black neighborhoods according to the Chicago Sun-Times.¹³ Since starting this program, Pappas has returned over \$108 million to homeowners ignorant of eligible refunds according to the Treasurer's Office.¹⁴ Pappas' initiative highlights racial disparities and inequity in property taxation while reestablishing trust among residents exploited by the complex bureaucratic system. Pappas hosts this segment weekly on the radio station WVON to answer questions via phone from homeowners and aid them in finding refunds and missing property tax exemptions according to the WVON Radio Outreach Segment.¹⁵ Publicly exposing gaps in the bureaucracy and a lack of transparency fosters public trust in public officials by making them aware of tax exemptions that benefit their finances as opposed to the government's.

In addition, Treasurer Maria Pappas attends many public events daily. One event Pappas appeared at was a public exhibition called "Trapped in the Shadows," an art display created by a Romanian painter and visual artist Carmen Marin on October 14th, 2025, in the Treasurer's

¹¹ "Cook County Homeowner Relief Fund." Cook County, accessed November 29, 2025.

¹² Ibid.

¹³ Chicago Sun-Times. "Black Houses Matter: Maria Pappas Radio Jock Feature." 2024.

¹⁴ Cook County Treasurer's Office. "Black Houses Matter Initiative Refund Announcement." n.d.

¹⁵ Cook County Treasurer's Office. "WVON Radio Outreach Segment." n.d.

Office in Chicago according to the Treasurer's Office.¹⁶ The Consul General of Romania joined Pappas in Chicago for the art exhibition. This cultural art event showed the "strong ties between Chicago and Romania and supports the ongoing efforts of Treasurer Pappas to promote diverse global artists within Cook County."¹⁷ Beyond trying to combat inequity in Latino and Black neighborhoods in Chicago, by trying to increase the diversity of culture in Chicago and retaining its relationship with Romania, Pappas promotes public trust in public officials of Cook County and therefore in the complexity of the government beyond its bureaucratic structure.

On the other hand, manual school closures in Cook County contribute to the complexity of the Cook County bureaucracy and hindrance of public trust. In Cook County, the Schiller Park school district states "The Superintendent of Schools is authorized by the Board of Education's policy to close schools on an emergency basis" according to the Schiller Park School District 81 on Emergency School Closing Information¹⁸ Similarly, the Palatine-Rolling Meadows school district states that "the school district takes the responsibility of deciding whether to close schools seriously" according to CCSD15 Palatine-Rolling Meadows on Emergency School Closings.¹⁹ Although this might appear as directed toward dangerous weather conditions, in 2013, the Chicago Board of Education shut down 50 public schools manually in mostly underdeveloped neighborhoods and minority communities causing large amounts of resentment and distrust of community members in the Cook County bureaucracy, and these shutdowns were not correlated to snow days. These sudden manual school closures underlined inequity in city funding and planning causing increased distrust towards the government according to WTTW News.²⁰

As a result, voter turnout in communities mostly affected by the bureaucracy of Cook County, through various ways including high property taxes and manual school closures, had the lowest voter turnout according to a study conducted in 2024 by the Cook County Treasurer. Lower income neighborhoods (mostly affected) who had a median household incomes of \$25-50k had an average voter turnout of 44.9%; whereas wealthy neighborhoods, rarely affected by the complexity of Cook County's bureaucracy, who have a median household income of

¹⁶ Cook County Treasurer's Office. "Trapped in the Shadows Art Exhibition." n.d.

¹⁷ Ibid.

¹⁸ Schiller Park School District 81. "Emergency School Closing Information." n.d.

¹⁹ CCSD15 Palatine-Rolling Meadows. "Emergency School Closings." n.d.

²⁰ WTTW News. "Monday Marks 10 Years Since Vote to Close 50 Chicago Public Schools." 2023.

\$150k+ had the highest voter turnout recorded in 2024 of 62.9%.²¹ According to this study, the Treasurer’s analysis found that “voters in the lowest-income taxing districts tended to complete ballots less often than voters in the wealthier taxing districts. Majority Black, Latino, or majority-minority populations tend to make up lower-income taxing districts, and the analysis of voting data highlights this correlation” according to the Treasurer’s Office on Voter Turnout 2024 Analysis.²²

In support of this finding, previous Director of Media and Public Relations for the Cook County Treasurer’s Office Ted Slowik provided his insight into the complexity of property taxes in Cook County. In an interview, Slowik said, “Cook County issues property taxes for over 2,200 taxing bodies (government bodies that have the legal power to collect and impose taxes). There’s a great deal of inequity in Cook County, throughout Illinois, and nationwide. Today, especially in the south suburbs – predominantly Black communities – you see extremely low collection rates (low numbers of citizens paying their taxes fully). In towns like Harvey, the collection rate has fallen to 50% or below. Meanwhile, places like Lake Forest have collection rates around 98–99%. Countywide, that’s about average, but these struggling communities are in a “death spiral.”” Slowik explained that the “death spiral” came into effect when the “industry left – the steel mills and factories closed, so those major employers stopped paying large portions of the tax base. Then the people who worked in those mills lost their jobs and couldn’t pay property taxes. Many left, and others abandoned their homes. Houses burned down or were demolished, turning into vacant lots that generate no tax revenue. That leaves fewer remaining residents to shoulder the entire tax burden. Rates skyrocket. No one wants to move there; no business wants to invest because taxes are so high. This cycle keeps repeating.” Inequity in Cook County’s property taxes in minority neighborhoods directly connects to public trust according to Slowik, “Transparency and accountability are always the most important factors. Property taxes are complicated, and most people don’t understand the system. The Treasurer does a good job educating the public on levies and how governments collect revenue – but no one attends the public hearings where tax levies are decided. Civic engagement at that critical moment is basically nonexistent.”

²¹ Cook County Treasurer’s Office. “Voter Turnout 2024 Analysis.” n.d.

²² Ibid.

As voter turnout in Cook County has decreased, especially in Black and Latino neighborhoods in the south suburbs, where property taxes have heavily increased, inequity in property taxes can be directly tied to a decline in voter turnout. Furthermore, lower appeal rates worsen inequity and contribute to a loss in voter turnout because when less appeals are being made, inequity increases, housing becomes more unaffordable, and trust in Cook County's bureaucratic system is lost – thus causing less people affected to vote or participate in Cook County's government. However, the Cook County Assessor's Homeowner Relief Fund helps to improve this issue by facilitating homeowner's financial struggles regarding property taxes specifically in the south suburbs. Cook County's Assessor's relief initiative increases awareness of property tax rights and the processes to refund overpaid property taxes among Black and Latino communities. Furthermore, "Black Houses Matter" augments the public's view of equality and responsiveness from the Cook County Treasurer. Pappas models how proactive communication and direct outreach toward minority communities in Cook County can rebuild public trust in the local government and its complex bureaucratic system. Despite Pappas' efforts to restore equity and public trust in Cook County's government, other governmental actions such as Chicago's Board of Education shutting down 50 public schools and more manual school closure repudiate public trust efforts and accelerate inequity in Cook County as the south suburbs homeowners' children get displaced.

Interestingly, although Cook County's public administration and bureaucracy can best be described as following Weber's bureaucratic rational/legal model or the Weber Theory, elements of this Theory can be proved false in the bureaucracy. Cook County does use those in professional positions (officials) to conduct political policies and bureaucratic actions and its authority does not lie in the hands of one central figure head or tradition; however, since Cook County's authority is not solely reliant on legal rules or regulations, as the many manual school closures and inequity in allocating different property taxes on certain neighborhoods are not rules to be administered but more so actions created by certain officials with different agendas, the Weber Theory cannot be used to fully describe Cook County's bureaucratic system. As the Weber Theory continues to be inapplicable to Cook County, public trust and equity decrease.

In conclusion, voter turnout in Cook County has decreased and is the lowest in low-income minority neighborhoods in the south suburbs, property taxes have increased and

appeal rates are much lower for people of color in Cook County, public schools were manually shut down, Treasurer Maria Pappas has been involved in outreach and public trust agendas, and communities affected by increased property taxes are in a “death spiral” according to previous Cook County official Ted Slowik.

Cook County’s complex bureaucracy mostly hinders public trust in local governance. While the Assessor’s Homeowner Relief Fund and outreach efforts by one county official aim to address inequity, the three-year reassessment cycle and uncoordinated network of smaller governments and departments sustain property-tax inequities, lower appeal rates in minority neighborhoods, and the manual shutdown of many Chicago public schools. This instability fuels mistrust and broadens voter-turnout gaps between wealthier neighborhoods protected by the bureaucracy and the minority, substandard neighborhoods it harms.

The complexity of Cook County’s bureaucracy and shrinking correlation to the Weber Theory reveal aspects of its government that hinder public trust and lower voter turnout in the minority, poorer neighborhoods that it affects. Repetitive patterns of these neighborhoods and homeowners suffering financially can be tied to the incoordination and inequity caused by Cook County’s bureaucracy.

Albeit arduous, viable solutions to making Cook County’s bureaucracy operate more efficiently in the future include the “Need for more civics education in schools and more effort from nonprofits – like the League of Women Voters – to teach transparency, accountability, and basic understanding of property taxes” according to previous county official Ted Slowik. If no changes are made, however, Cook County property taxes will worsen. Slowik said, “Infrastructure needs are enormous. Water systems in communities like Park Forest are deteriorating. These towns can’t afford upgrades. They look to the federal government for major assistance. The best long-term solution is economic development. Amazon, for instance, tore down dilapidated housing in Markham and built a distribution center. That’s good – jobs, new economic activity.” Therefore, economic development stabilizes communities and improves the tax base over time. More importantly, to sustain economic growth, long-term county level solutions need to be made such as “economic incentives for development and targeted investment in underserved communities. It is beneficial to funnel infrastructure dollars to areas that have historically been neglected – repairing roads, improving parks. That’s the right move.”

In order to make Cook County’s bureaucracy more efficient and foster public trust while growing the economy, long-term county-level solutions need to be put into action and minority neighborhoods repeatedly targeted by the bureaucracy need easily accessible and free economic relief so that they can live healthily and happily in their homes in Cook County.

Works Cited

Cook County Assessor’s Office. “About Cook County Assessor’s Office.” n.d.
<https://www.cookcountyassessoril.gov/about-cook-county-assessors-office>

Cook County Government. “About Cook County.” n.d.
<https://www.cookcountyil.gov/about-cook-county>

“Cook County Homeowner Relief Fund.” Cook County, accessed November 29, 2025.
<https://www.cookcountyil.gov/service/cook-county-homeowner-relief-fund>

Cook County Treasurer’s Office. “Duties and Responsibilities.” n.d.
<https://www.cookcountytreasurer.com/dutiesandresponsibilities.aspx>

Cook County Treasurer’s Office. “Black Houses Matter Initiative Refund Announcement.” n.d.
<https://www.cookcountytreasurer.com/newsarticle.aspx?articleid=1497>

Cook County Treasurer’s Office. “WVON Radio Outreach Segment.” n.d.
<https://www.cookcountytreasurer.com/newsarticle.aspx?articleid=1014>

Cook County Treasurer’s Office. “The \$2 Billion Wealth Transfer: How Cook County’s Property Tax Appeals System Rewards Businesses at the Expense of Homeowners.” *Cook County Treasurer’s Office*, May 5, 2025. <https://cookcountytreasurer.com/newsarticle.aspx?articleid=2090>

Cook County Treasurer’s Office. “Voter Turnout 2024 Analysis.” n.d.
<https://www.cookcountytreasurer.com/pdfs/voteturnout/VoterTurnout2024AnalysisEnglish.pdf>

Cook County Treasurer’s Office. “Trapped in the Shadows Art Exhibition.” n.d.
<https://www.cookcountytreasurer.com/newsarticle.aspx?articleid=2147>

Investopedia. “Property Tax.” n.d.
<https://www.investopedia.com/terms/p/propertytax.asp>

Lake County, Illinois. “Appeal Process.” n.d.
<https://www.lakecountyil.gov/503/Appeal-Process>

Plato Stanford Encyclopedia. "Republicanism." n.d.
<https://plato.stanford.edu/entries/republicanism/>

Schiller Park School District 81. "Emergency School Closing Information." n.d.
<https://www.sd81.org/about/emergency-school-closing-information>

CCSD15 Palatine–Rolling Meadows. "Emergency School Closings." n.d.
<https://www.ccsd15.net/our-district/departments/administrative-services/school-student-safety/emergency-school-closings>

U.S. Census Bureau. "Glossary: Census Tract." n.d.
https://www.census.gov/programs-surveys/geography/about/glossary.html#par_textimage_13

WTTW News. "Monday Marks 10 Years Since Vote to Close 50 Chicago Public Schools." 2023.
<https://news.wttw.com/2023/05/22/monday-marks-10-years-controversial-vote-close-50-chicago-public-schools>

Chicago Sun-Times. "Black Houses Matter: Maria Pappas Radio Jock Feature." 2024.
<https://chicago.suntimes.com/columnists/2024/03/24/black-houses-matter-maria-pappas-radio-jock-cook-county-treasurer-wvon>

South Side Drive Magazine. "Black Houses Matter: Mission to Save Your Property." n.d.
<https://southsidedrivemag.com/black-houses-matter-treasurer-maria-pappas-mission-to-save-your-property-by-emma-young/>

Axios Chicago. Reporting on Cook County Tax Assessments. n.d.
<https://www.axios.com/local/chicago>

WBEZ Chicago. Reporting on Cook County Reassessments. n.d.
<https://www.wbez.org>

Index

Parcels: Specific, defined pieces of property recognized as distinct entities for legal ownership, land use, and taxation.

Residential tax bills: A statement detailing the payment property owners must pay to fund government departments and projects.

County relief program: County-funded program that provides cash aid if you are an adult without any income or resources, and children in certain special circumstances who are ineligible for federal or State programs.

Property tax: A tax paid by property owners, based on the value of land and buildings they own. Local governments typically use it to fund public services like schools, police, and infrastructure (Investopedia).

Bureaucracy: A system of government or organization that uses standardized rules and procedures, hierarchical authority, and specialized roles to manage complex operations (Britannica, "*Bureaucracy*").

Administration: The process or group responsible for managing and directing the operations of an organization or government (Merriam-Webster, "*Administration*").

Weber's rational-legal model: Sociologist Max Weber's concept that authority in modern governments is based on legal rules and procedures, not personal loyalty or tradition. Power comes from laws and professional positions (Britannica, "*Rational-Legal Authority*").

Municipality: A city, town, or local administrative district that has its own government and provides local services (Britannica, "*Municipality*").

Republicanism: A political philosophy emphasizing civic participation, rule by elected representatives, and the public good, rather than rule by a monarch (Stanford Encyclopedia of Philosophy, "*Republicanism*").

Public trust: The belief that government officials are acting in the public's best interests, fairly and honestly (Britannica, "*Public Trust Doctrine*").

Inequity: Unfairness or unequal treatment, especially when resources, rights, or opportunities are distributed unevenly (Merriam-Webster, "*Inequity*").

Turnout: The percentage of eligible voters who actually cast a vote in an election (Britannica, "*Voter Turnout*").

Appeal: A request for a higher authority (such as a court or governmental body) to review and change a decision made by a lower authority (Cornell Law School, Legal Information Institute, "*Appeal*").

Cook County Treasurer: In Cook County, the Treasurer oversees the property-tax collection and distribution system. Duties include printing and mailing bills; collecting \$16.1 billion annually from more than 1.8 million parcels; and distributing funds to approximately 2,200 local government agencies (Cook County Treasurer's Office).

Cook County Assessor: The elected official responsible for establishing fair and accurate property assessments for the county's 1.8 million parcels for ad valorem tax purposes (Cook County Assessor's Office).

Census tract: Small, relatively permanent statistical subdivisions of a county used by the Census Bureau for data collection (U.S. Census Bureau, "*Census Tracts*").

Utilising pomelo peel (*Citrus grandis*) as a sustainable bio-absorbent to tackle water scarcity: A novel agricultural research by Anwesha Shantanu Choudhari, Rajas Raghav Khare, Sitalakshmi Parameswaran, Revathi Haval, Sheetal Shantanu Choudhari, and Vibhuti Shreesh Mhatre.

Abstract

Pomelo peel, often regarded as an agricultural waste, can serve as a potent natural bio-absorbent. The present study explored the use of Pomelo albedo as a bio-absorbent to conserve water and improve soil health. The results of our study demonstrated its effectiveness in retaining water and its ability to absorb oil-based impurities, suggesting it could be a valuable tool for improving soil quality and combating water scarcity. However, further research is needed to evaluate its long-term effectiveness, scalability, and environmental impact in real-world agricultural contexts.

Keywords: Pomelo albedo, *Citrus Grandis*, Bio-absorbent, Sustainable agriculture, Agricultural soil management.

Introdcution

Water scarcity is becoming a global crisis, threatening food security due to its direct impact on agriculture. As crops depend on consistent water access, diminished availability increases the risk to food production. Traditional soil and water management systems have often proven too costly, necessitating the exploration of sustainable, low-cost alternatives. One promising solution lies in the natural properties of Pomelo peel (*Citrus grandis*), particularly its albedo. Pomelo peel, often regarded as agricultural waste, is rich in cellulose and can serve as a potent natural bio-absorbent. Cellulose-based materials, such as Pomelo peel, are effective in absorbing water and oil, which makes them valuable for addressing issues related to water retention and soil quality in agriculture.[1]

Pomelo peel's pectin, essential oils, and active polysaccharides give it remarkable water absorption properties, making it ideal for improving soil fertility and reducing erosion.[2,3] Research on cellulose-based aerogels has shown that natural fibres can efficiently absorb oil and water, providing a sustainable approach to managing oil-contaminated water.[4] The potential for Pomelo peel to enhance soil quality by acting as an absorbent material is significant, especially in tropical regions like India, where highly acidic soils and over-exploitation have led to reduced fertility.[5]

Materials and methods

The present study explores the use of Pomelo albedo as a bio-absorbent to conserve water and improve soil health. In the first experiment, various-sized dried Pomelo albedo pieces were submerged in water and soil to test water absorption and retention capacity. The second experiment focused on its oil-absorption capacity in an oil-water mixture, as efficient oil absorption is important for nutrient retention and maintaining soil fertility. The results of this research could provide farmers with a sustainable, biodegradable alternative to chemical water retention solutions.[6]

In this study, the potential of Pomelo albedo (*Citrus Grandis*) as a natural bio-absorbent was evaluated through two separate experiments. The albedo, a thick spongy layer from Pomelo fruit, which was taken from Afjan. was dried and divided into various sizes for testing. The first experiment aimed to assess the water absorption and retention capacity of different-sized Pomelo albedo pieces. Six sets of Pomelo albedo, weighing 5 grams (two sets), 8 grams, 10 grams, 14 grams, and 15 grams (three pieces per set), were used. These pieces were submerged in three different mediums: 250 ml of water, 250 grams of wet soil, and 250 grams of dry soil. The samples were left in the mediums for 24, 48, and 72 hours, during which time their weights were measured using a calibrated digital weighing balance. The samples were weighed before submersion and at the specified time intervals to observe any changes in weight. This allowed for the calculation of the water absorption capacity of the Pomelo albedo under different conditions. All samples were maintained under similar environmental conditions to ensure consistency.

The second experiment focused on testing Pomelo albedo's ability to absorb oil-based impurities and its potential to improve soil quality. For this, different-sized Pomelo albedo pieces

were submerged in an oil-water mixture consisting of 100 ml of oil and 150 ml of water. To visually observe oil absorption, red food colouring was added to the mixture. The Pomelo pieces were submerged for 12 hours. Their weights were recorded before and after the submersion to evaluate the absorption of oil. By analysing the weight difference, the efficiency of Pomelo albedo in removing oil-based impurities was assessed. Both experiments provided valuable insights into the bio-absorbent properties of Pomelo albedo and its potential for agricultural applications.[7]

Results

Experiment 1- % absorption/retention was calculated using the formula:

% Absorption/Retention=

Weight of sample after submerged - Weight of the sample before submerged /Weight of sample before x 100

Table 1.1 shows an increase in the weight of pomelo albedo after being submerged in water for 24, 48, and 72 hrs, irrespective of its size. Tables 1.2 a, b & c show the average water absorption and retention for small, medium, and large pieces of pomelo albedo. There was also an increase in the weight of the pomelo albedo when placed in wet soil, and water retention was found to change with the change in size of the pomelo albedo. The absorption capacity of large pieces is highest, while the retention capacity is highest for small pieces. It was observed that after being submerged in dry soil, there was no decrease in the weight of pomelo albedo.

Size of Pomelo	Before Experiment (Wt in grams)	After 24 hours (Wt in Grams)			After 48 hours (Wt in Grams)			After 72 hours (Wt in Grams)		
		Water	Wet Soil	Dry Soil	Water	Wet Soil	Dry Soil	Water	Wet Soil	Dry Soil
Small	5.00	6.00	8.00	5.00	7.00	9.00	6.00	7.00	8.00	5.00
	5.00	5.30	5.19	5.05	6.83	5.68	5.70	7.42	5.70	5.05
Medium	8.00	11.00	9.00	8.00	13.00	10.00	7.00	11.00	9.00	8.00
	10.00	12.49	11.87	10.20	11.71	13.10	10.08	11.91	12.69	10.01
Large	14.00	21.00	17.00	14.00	20.00	16.00	14.00	23.00	17.00	14.00
	15.00	20.50	19.50	15.40	20.30	17.23	15.60	20.44	17.74	15.61

Table 1.1: Change in weight of different-sized Pomelo albedo after 24 hrs, 48 hrs, and 72 hrs after being submerged in water, wet and dry soil.

Table 1.2 a, b & c shows % average Absorption/ Retention after 24, 48 & 72 hrs, respectively.

Size of Pomelo	After 24 hrs	
	Average Absorption	Average Retention
Small	13.00 %	31.90 %
Medium	31.20 %	15.60 %
Large	43.33 %	25.71 %

Table 1.2a- % % Average Absorption/ Retention after 24 hrs.

Size of Pomelo	After 48 hrs	
	Average Absorption	Average Retention
Small	38.30 %	46.80 %
Medium	39.80 %	28.00 %
Large	39.10 %	14.58 %

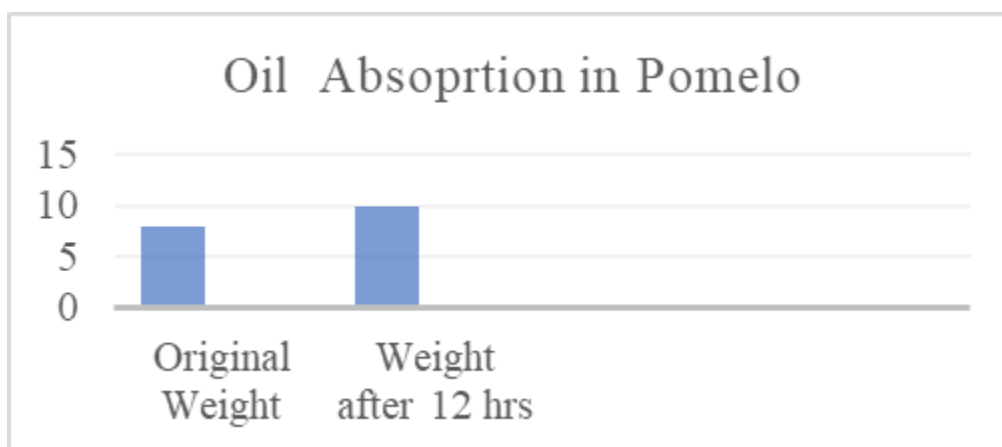
Table 1.2b- % Average Absorption/ Retention after 48 hrs.

Size of Pomelo	After 72 hrs	
	Average Absorption	Average Retention
Small	44.20 %	37.00 %
Medium	28.30 %	19.70 %
Large	50.28 %	19.85 %

Table 1.2c - % Average Absorption/ Retention after 72 hrs

	Original Weight (gm)	Weight after 12 hrs (gm)
Oil	8 gm	10 gm

Table 2- Oil absorption of Pomelo



Graph 1- Oil absorption in Pomelo

Experiment 2: Different-sized pomelo albedo pieces were submerged in the oil-water mixture for 12 hours. Readings of the second experiment showed that Pomelo albedo had turned red (food colour) and its weight increased, which shows that Pomelo albedo can absorb oil. (Table 2)

Discussion

The current study aimed to investigate the potential of Pomelo albedo (*Citrus grandis*) as a natural bio-absorbent for enhancing water retention in soil and removing oil-based impurities, to address issues related to water scarcity and soil quality. The results from the experiments supported the hypothesis that Pomelo albedo is an effective natural bio-absorbent, providing promising insights into its potential for agricultural applications. The present study explores the implications of Pomelo albedo for sustainable agriculture.

The first experiment, which tested the water absorption capacity of Pomelo albedo across different sizes, demonstrated significant water retention over 24, 48, and 72 hours. Larger-sized pieces, particularly those weighing 15 grams, absorbed and retained the most water. This shows the water absorption ability of pomelo albedo, confirming the hypothesis that size plays a key role in the absorption capacity. There was no change in the weight of Pomelo Albedo in the dry soil. This shows that pomelo albedo can hold and retain water in dry soil. This shows the water retention ability of pomelo albedo. This result aligns with previous research by Wang et al. (2024) [1], who demonstrated that Pomelo peels, due to their high cellulose content, act as

efficient sorbents for water. The high cellulose concentration in Pomelo albedo enables it to form a sponge-like structure, making it highly effective in retaining moisture.

In the second experiment, which aimed to test Pomelo albedo's ability to absorb oil-based impurities, the results were equally promising. After submerging the albedo in an oil-water mixture for 12 hours, it was evident that Pomelo albedo absorbed a considerable amount of oil, leaving the water less contaminated. The addition of red food colouring made it clear that oil, rather than water, was absorbed, indicating Pomelo albedo's hydrophobic properties in this scenario. This finding resonates with the work of Shi et al. (2019) [4], who found that cellulose-based materials, such as Pomelo peel-based aerogels, are excellent for absorbing oil and other organic pollutants. The result is particularly relevant for improving soil quality, especially in regions where oil-based pollutants might threaten the health of agricultural lands.

One of the key strengths of this study is the dual benefit of using Pomelo albedo for both water retention and pollutant removal. As water scarcity becomes more severe due to global climate change, farmers need sustainable, cost-effective solutions to retain soil moisture and reduce reliance on chemical additives. The natural properties of Pomelo albedo, particularly its cellulose content and biodegradable nature, make it a promising candidate for soil improvement and water conservation. In this study, the biodegradable nature of Pomelo albedo was considered beneficial, as it can be used without harming the environment. Once integrated into the soil, the material would break down naturally, providing nutrients to the soil and potentially enhancing its overall quality.

Comparing these findings to other studies highlights the uniqueness of Pomelo albedo as a multifunctional bio-absorbent. Previous research on cellulose-based materials like cotton, kapok fibre, and banana peels has shown promising results in terms of water retention and pollutant absorption. [2,3], but Pomelo albedo offers an advantage due to its relatively high availability and low cost. As noted by Visakh et al. (2022) [7], citrus peel waste is an abundant byproduct of the fruit industry, and utilizing it for agricultural purposes would help reduce environmental waste while providing an economic and ecological benefit.

While the results are encouraging, certain limitations must be acknowledged. First, the experiments were conducted under controlled laboratory conditions, which may not fully replicate the challenges present in a real agricultural environment. Factors such as varying temperatures, soil compositions, and exposure to natural elements could influence the effectiveness of Pomelo albedo in field applications. Moreover, while the study focused on short-term water retention and pollutant absorption, the long-term effects of incorporating Pomelo albedo into the soil were not explored. Future research should investigate the decomposition rate of Pomelo albedo in different soil types and climates to assess its durability and overall impact on soil health.

Another aspect worth exploring in future studies is the scalability of this method. While Pomelo peel is a common byproduct of citrus fruit production, large-scale applications of this material in agriculture would require significant volumes of peel, which might pose logistical challenges. Investigating how to process and store Pomelo albedo for widespread agricultural

use, perhaps by converting it into more easily distributable forms like powders or pellets, could make it more feasible for farmers to adopt this method.

In conclusion, this study confirms that Pomelo albedo has substantial potential as a natural bio-absorbent for water retention and pollutant removal in agricultural settings. Its effectiveness in retaining water, coupled with its ability to absorb oil-based impurities, suggests that it could be a valuable tool for improving soil quality and combating water scarcity. However, further research is needed to evaluate its long-term effectiveness, scalability, and environmental impact in real-world agricultural contexts. If proven feasible, the use of Pomelo albedo could represent a significant step forward in sustainable agriculture, offering a green alternative to chemical soil amendments and contributing to the global effort to combat water scarcity.

Works Cited

1. Wang, H., Wang, P., Kasapis, S., & Truong, T. (2024). Pomelo (*Citrus grandis* L.) peels as effective sorbents for diverse gel matrices: The influence of particle size and powder concentration. *Journal of Food Engineering*, 370, 111966. <https://doi.org/10.1016/j.jfoodeng.2024.111966>
2. Huang, C. Y., Liu, X. Z., & Wu, J. P. (2014). Pectin, flavonoids, and other bioactive components in Pomelo peel: Functional potential. *Journal of Agricultural and Food Chemistry*, 65(2), 234-245.
3. Liu Y, Liu A, Ibrahim SA, Yang H, Huang W. Isolation and characterization of microcrystalline cellulose from pomelo peel. *International Journal of Biological Macromolecules*. 2018 May 1;111:717-21.
4. Shi, G., Qian, Y., Tan, F., Cai, W., Li, Y., & Cao, Y. (2019). Controllable synthesis of pomelo peel-based aerogel and its application in adsorption of oil/organic pollutants. *Royal Society Open Science*, 6(2). <https://doi.org/10.1098/rsos.181823>.
5. Mandal D, Mandal C, Prasad J, Bhattacharyya T, Acid Soils in Agro-Ecological Sub-regions of India: A Revisit. *Indian Journal of Fertilisers* 2019;15(10):1156-1166.
6. Gyawali D, Rijal S, Basnet P, Ghimire KN, Pokhrel MR, Paudyal H. Effective biosorption of As(V) from polluted water using Fe(III)-modified Pomelo (*Citrus maxima*) peel: A batch, column, and thermodynamic study. *Heliyon*. 2023 Feb 3;9(2):e13465. doi: 10.1016/j.heliyon.2023.e13465. PMID: 36816270; PMCID: PMC9929298.
7. Visakh NU, Pathrose B, Narayanankutty A, Alfarhan A, Ramesh V. Utilization of Pomelo (*Citrus maxima*) Peel Waste into Bioactive Essential Oils: Chemical Composition and Insecticidal Properties. *Insects*. 2022; 13(5):480. <https://doi.org/10.3390/insects13050480>.

Gut Microbiome By Rishi Jain

In recent years, knowledge about the human gut microbiome has grown due to modern technological advancements ([Shreiner, Kao, and Young 2015](#)). This microbiome houses trillions of microbes—roughly three times the number of human cells in the body ([Ferranti et al. 2014](#), [Roy and Conroy 2018](#)). It contains over 300 bacterial species, many of which are essential for activities like fighting harmful pathogens and the metabolism of nutrients ([Quigley 2013](#), [Rinninella et al. 2019](#)). What makes this microbiome particularly interesting is its ability to connect cognitive and emotional parts of the brain through peripheral intestinal abilities via the gut-brain axis ([Carabotti et al. 2015](#)). This axis refers to the ongoing exchange of information between the brain (central nervous system) and the gut's enteric nervous system ([Carabotti et al. 2015](#)). It operates through many different pathways including the immune system and the vagus nerve. ([Cryan et al. 2019](#)). Disruption of this system, often due to microbial imbalance (dysbiosis), has been linked to many neurodegenerative diseases like Alzheimer's disease (AD) and Parkinson's disease (PD) ([Borrego-Ruiz and Borrego 2024](#)). These conditions share a common feature of neuron loss in the central nervous system. ([Gao and Hong 2008](#)). Interestingly, growing evidence demonstrates that the gut microbiome composition differs significantly between individuals with these neurodegenerative diseases when compared to healthy individuals. ([Ma et al. 2024](#)). For example, patients with amyotrophic lateral sclerosis (ALS) had a reduced presence of butyrate-producing bacteria like *Eubacterium rectale* and *Roseburia intestinalis*, when compared to their healthy counterparts ([Chen et al. 2024](#)). These bacteria are essential for maintaining intestinal barrier integrity ([Picón-Jaimes et al. 2023](#)). Similarly, Parkinson's disease has been characterized by a decline in a *Prevotella* bacterial species and *Eubacterium rectale*, which have been associated with cognitive decline ([Chandra et al. 2023](#), [Wegrzyn et al. 2024](#)). These studies demonstrate that microbial alterations may play a role in the development or trajectory of neurodegenerative diseases ([Zhang et al. 2024](#)). Importantly, the composition of the gut microbiome can continuously change— it responds to environmental and behavioral factors, including antibiotics, stress, and lifestyle ([Rinninella et al. 2019](#)). Diet is one of the most influential modulators ([Rinninella et al. 2019](#)). For example, fermented foods (such as yogurt, kombucha, and kefir) have the potential to introduce beneficial bacterial species like *Lactobacillus* and *Bifidobacterium* to the human gut ([Valentino et al. 2024](#), [Skowron et al. 2022](#)). *Lactobacillus* is able to stimulate beneficial lactic acid production ([Chandrasekaran, Weiskirchen, and Weiskirchen 2024](#)). *Bifidobacterium*, has been shown to prevent memory impairment in mice with neurodegenerative diseases ([Abdelhamid et al. 2022](#)). Exercise, another potential therapeutic avenue, exerts similar effects. Namely, it can increase the presence of certain beneficial bacteria, namely *Prevotella*, *Faecalibacterium prausnitzii* ([Rojas-Valverde 2023](#)). For one thing, *Prevotella* has been proven to improve glucose metabolism ([Kovatcheva-Datchary et al. 2015](#)). Meanwhile, *Faecalibacterium prausnitzii*, has been correlated with an increase in cognitive function for Alzheimer's disease patients ([Ueda et al. 2021](#)). Together, all of these findings show the potential for lifestyle

interventions—specifically, fermented food consumption and regular exercise—to beneficially modulate the gut microbiome and support cognitive health ([Kovatcheva-Datchary et al., 2015](#); [Ueda et al., 2021](#); [Clarke et al., 2014](#)). By targeting microbial diversity, metabolic function, and gut–brain communication, these strategies may help slow the progression of neurodegenerative diseases. This review therefore examines the current evidence on how diet and exercise shape the gut–brain axis, evaluates their impact on the human microbiome, and neurodegeneration, and explains how these techniques can be used to deal with neurodegenerative diseases.

Fermented foods are produced through microbial growth ([source](#)). They usually contain some form of live bacterial organisms, many of which are beneficial for the body, and have the ability to modulate the brain-gut axis through different ways including with the immune system .. ([source](#)) ([source](#)). In a 12-week randomized and controlled trial, adults aged 65 and older were asked to consume a fermented soy powder without probiotics (added to their daily diet), while a control group was provided a placebo ([PubMed](#), [MDPI](#)). Cognitive features such as memory, verbal fluency, processing speed, and executive function were measured before and after that intervention ([PubMed](#)). The group receiving fermented soy showed significant improvements in memory performance, ([MDPI](#)), although no significant changes were observed in the control group ([PubMed](#)). Senescence-accelerated mouse prone 8 (SAMP8) mice were given daily doses of natto, a fermented soybean product, for a period of 12 weeks ([RSC Publishing](#)). Spatial learning and memory were assessed through tests involving behavior, ([RSC Publishing](#)). Hippocampal molecular signaling pathways, including BDNF, CREB, TrkB, and NMDAR-CaMKII, were analyzed ([RSC Publishing](#)). The mice fed natto showed improvements in spatial learning and memory ([RSC Publishing](#)) and showed activation of important neurotrophic signaling pathways in the hippocampus ([RSC Publishing](#)). Male mice that had memory problems from scopolamine, and female mice that were ovariectomized and given D-galactose to mimic aging, were treated with fermented ginseng each day at different doses ([MDPI study 1](#)). Their behavior was tested using the Morris Water Maze ([MDPI study 1](#)), and researchers measured things like acetylcholinesterase (AChE) activity, oxidative stress levels such as malondialdehyde (MDA), and antioxidant enzymes like catalase and superoxide dismutase ([MDPI study 2](#)). The fermented ginseng improved spatial memory, lowered AChE activity, reduced oxidative stress, and boosted antioxidant enzyme activity ([MDPI study 2](#)). In the NHANES 2011–2014 cohort, which included approximately 2,462 adults with an average age of 69 years, dietary intake of fermented dairy products such as yogurt and cheese was assessed through self-reported dietary recall ([PubMed](#)). Cognitive performance was measured using standardized tests including the Animal Fluency Test, Digit Symbol Substitution Test, and composite global cognitive scores ([PubMed](#)). Participants with higher yogurt consumption exhibited better global cognition and verbal fluency compared to those with low or no fermented dairy intake ([PubMed](#)). Older adults aged 60–95 years diagnosed with Alzheimer’s disease were assigned to receive a probiotic-enriched milk, containing strains such as *Lactobacillus acidophilus*, *L. casei*, *L. fermentum*, and *Bifidobacterium bifidum*, for 12 weeks ([MDPI](#)). The control group received non-probiotic milk ([MDPI](#)). Cognitive function was assessed using the

Mini-Mental State Examination (MMSE) ([MDPI](#)), and blood biomarkers related to inflammation (hs-CRP), oxidative stress (malondialdehyde), and metabolic status (insulin resistance) were measured ([MDPI](#)). Patients receiving the probiotic milk experienced significant improvements in MMSE scores ([MDPI](#)), reduced oxidative stress ([MDPI](#)), lower inflammatory markers ([MDPI](#)), and improved metabolic parameters ([MDPI](#)).

A systematic review and meta-analysis of 22 randomized controlled trials, including 1,551 participants, examined the effects of fermented foods, prebiotics, and probiotics on cognitive outcomes ([PubMed](#)). Six of the included studies, involving approximately 472 participants, specifically investigated fermented food interventions ([PubMed](#)). Outcomes analyzed included global cognition and cognitive subdomains such as memory, attention, and executive function ([PubMed](#)). While several individual studies reported improvements in specific cognitive domains, the overall pooled analysis did not find statistically significant effects of fermented food interventions on global cognitive function ([PubMed](#)). A 2023 systematic review evaluated epidemiological evidence on the relationship between fermented food and beverage consumption—including fermented soy products—and cognitive decline in older adults ([PubMed](#)). The review found that higher daily intake of fermented foods was associated with a slower rate of cognitive decline ([PubMed](#)) and a lower risk of developing dementia or Alzheimer's disease ([PubMed](#)). Several preclinical studies on rodents reviewed in a recent publication administered different fermented foods—such as camembert cheese extract, fermented soymilk containing *Lactobacillus plantarum*, Cheonggukjang, and red mould rice—to animals with induced Alzheimer's disease or cognitive impairment ([PMC](#)). Outcomes measured included A β deposition, synaptic markers (such as BDNF, GDNF), inflammatory cytokines, memory tasks, and oxidative stress markers ([PMC](#)). Fermented food supplementation consistently reduced A β deposition ([PMC](#)), improved synaptic plasticity markers ([PMC](#)), lowered inflammatory responses ([PMC](#)), enhanced performance in memory tests ([PMC](#)), and improved oxidative stress profiles ([PMC](#)).

Fermented foods refer to ideas that are produced through microbial growth ([source](#)). They usually contain some form of live bacterial organisms, many of which are beneficial for the body, and have the potential to modulate the brain-gut axis through Different ways such as the immune system .. ([source](#)) ([source](#)). In a 2024 study, the long-term effects of the probiotic strain *Streptococcus thermophilus* MN-ZLW-002 on Alzheimer's disease (AD)-related pathology and behavior using a mice model were investigated (1). The population consisted of male APP/PS1 mice, a model that is similar to human amyloid-beta pathology, (a substance that plays a special role in Alzheimer's disease) ([source](#)) starting at 3 months of age—a time window during which early amyloid deposition and behavioral changes typically emerge. Wild-type C57BL/6J mice were used as healthy controls in this experiment. The intervention involved daily oral administration of either live or heat-inactivated *S. thermophilus* MN-ZLW-002 for 12 weeks. Two dosages were tested: a low dose of 8.3×10^9 CFU/kg·day and a high dose of 1.67×10^{10} CFU/kg·day. After the 12-week treatment period, mice were observed for an additional 12 weeks without further supplementation, allowing for assessment of their post-treatment medical

condition. Researchers measured a wide range of outcomes. Cognitive performance was assessed through a Morris water maze, step-down passive avoidance, open field, and shuttle box tests. Amyloid- β levels were also measured in the hippocampus, along with markers of astrocyte activation (GFAP) and microglial activation (IBA-1). In addition, the researchers measured serum inflammatory markers (IL-1 β , IL-6, TNF- α , LPS, CRP), tight junction gene expression (occludin, ZO-1), short-chain fatty acids (notably propionate), and hippocampal neurotrophic and antioxidant markers like the brain-derived neurotrophic factor (BDNF), superoxide dismutase (SOD), catalase (CAT), and glutathione peroxidase (GSH-Px). The results were very strain-form dependent. However, it was discovered through the tests Both live and heat-inactivated forms of *S. thermophilus* improved spatial learning and memory in APP/PS1 mice. Notably, the heat-inactivated high-dose group showed a reductions in anxiety-like behavior, while the live high-dose group had a uniquely decreased hippocampal astrogliosis, potentially indicating a potential immunomodulatory role. However, none of the interventions changed hippocampal amyloid- β plaque levels, but both live and inactivated forms reduced systemic LPS levels—an indication, of endotoxemia—and increased intestinal tight junction gene expression, suggesting enhanced gut barrier integrity. The heat-inactivated form produced a significant rise in colonic propionic acid, a short-chain fatty acid known to show anti-inflammatory and neuroactive effects. Additionally, the group also showed increased BDNF expression and hippocampal SOD activity, maybe showing support for neuronal survival? From a neurodegenerative disease standpoint, these findings are important. The APP/PS1 model shows and recapitulates the early pathological and behavioral hallmarks of Alzheimer's disease, making it a good platform for testing interventions. That *S. thermophilus* MN-ZLW-002—especially in its heat-inactivated form—was able to produce lasting cognitive benefits after the treatment period ended, suggests a potential gut–brain axis modulation effect. The of amyloid pathology from behavioral improvement also reinforces the growing view that neuroinflammation, oxidative stress, and gut permeability are factors of disease progression. In terms of the topic of the review, this study directly supports that that modulation of the gut microbiota using specific bacterial strains can have protective effects on brain health, (especially in the presence of neurodegenerative pathology). It ALSO show the potential of targeting the gut–brain axis through precision bacterial interventions to improve cognitive outcomes and modulate inflammation. Importantly, it shows that both live and non-viable bacterial preparations can be effective, suggesting future strategies could be designed using more stable, heat-treated bacteria that are easier to store and give to people.

Taken this, this reinforces the concept that gut microbiome modulation is not merely correlative but may be causally linked to improved cognitive and neurological outcomes. It strengthens the case for using bacterial strains like *S. thermophilus* MN-ZLW-002 as a potential adjunct therapy in the management of Alzheimer's disease and related neurodegenerative conditions through the Gut-Brain axis.

Another study was conducted using 8 week old C57BL/6J mice (n=32), a widely used strain in neurodegenerative disease experiments due to its potential response to

neuroinflammatory problems. The goal of the study was to understand the effects of *Lactobacillus plantarum* on neuroinflammation and neurodegeneration, specifically in diseases such as Parkinson's disease (P) and Alzheimer's disease (AD). The mice were divided into four distinct groups. The Normal Control Group consisted of healthy mice that received no disease induction or treatment, establishing a baseline for comparison. The Control + Lac Group consisted of healthy mice treated with *Lactobacillus plantarum*, but without any disease induction. This group allowed the researchers to observe the effects of the probiotic in the absence of the disease. The Model Group received lipopolysaccharide (LPS) to induce neurodegeneration, representing a model of disease without any probiotic treatment. Finally, the Model + Lac Group received both the LPS-induced neurodegeneration and the *Lactobacillus plantarum* treatment, letting researchers test if the probiotic could mitigate the effects of neurodegeneration.

The intervention including administration of *Lactobacillus plantarum* to some selected groups. The treatment was administered over a period of 30 days, following the induction of neurodegeneration via LPS. This duration was chosen to discover the immediate and longer-term effects of the probiotic on neuroinflammation and behavioral outcomes. *Lactobacillus plantarum* was also selected for its known anti-inflammatory properties and its ability to modulate the gut microbiome. This strain has been shown to influence the gut-brain axis, a pathway thought to play a crucial role in neurodegenerative diseases, where disturbances in gut microbiota may contribute to neuroinflammation and neuronal damage. To assess the effects of the study, several key factors were measured. Inflammatory markers were one of the primary measurements. Specifically, the study focused on cytokines such as $\text{TNF-}\alpha$, $\text{IL-1}\beta$, and IL-6 , which are commonly analyzed during neuroinflammation and are associated with neurodegenerative diseases. The levels of these cytokines were measured both in the blood and brain tissue to determine the extent of systemic and central nervous system inflammation. Additionally, the activation of microglia, the primary immune cells in the brain, was assessed using markers such as Iba1 and CD11b, which can indicate microglial activation in response to injury or inflammation. In addition to inflammatory markers, behavioral outcomes were measured to assess the impact of the treatment on the mice's motor and cognitive function. This was done through tests such as the rotarod to evaluate motor coordination and balance, and memory tasks (e.g., the Morris water maze), which assess spatial learning and memory.

The primary outcome of the study was the assessment of inflammatory indicators in the Model + Lac Group compared to the Model Group, to determine if *Lactobacillus plantarum* could reduce the inflammatory response induced by LPS. The results showed that the probiotic significantly decreased the levels of $\text{TNF-}\alpha$, $\text{IL-1}\beta$, and IL-6 in both peripheral blood and brain tissue, indicating a marked reduction in neuroinflammation. Furthermore, the Model + Lac Group exhibited improved motor function as shown by their performance on a rotarod test, where they demonstrated better coordination and balance compared to the untreated Model Group. Similarly, cognitive performance, assessed through the Morris water maze, was significantly improved in the Model + Lac Group, suggesting that *Lactobacillus plantarum*

alleviates neuroinflammation and protects against the cognitive decline associated with neurodegeneration. Relation to Neurodegeneration and Overall Implications for the Thesis. The study was able to show that *Lactobacillus plantarum* holds a potential for mitigating neuroinflammation and protecting against neurodegenerative changes. In neurodegenerative diseases like Parkinson's disease (PD) and Alzheimer's disease (AD), inflammation plays a key role in accelerating neuronal damage and disease progression. The ability of *Lactobacillus plantarum* to reduce inflammatory cytokines and improve both motor and cognitive function is highly relevant for the development of adjunct therapies for these conditions. This evidence supports the hypothesis that gut microbiome modulation, particularly through probiotics like *Lactobacillus plantarum*, can represent a good strategy for reducing neuroinflammation and protecting against neuronal degeneration.

Housing Scarcity as an Institutional Outcome: Immigration, Capital Flows, and Urban Governance

By Matthew Lewis

Abstract

While immigration is often blamed for housing stress in European cities, supply does not respond in direct proportion to population change. Employment concentration in cities, labor and skills dispersion through employment markets, heterogeneous household income growth, and the movement of investment to places where returns are perceived to be less risky mean that housing supply does not readily adapt to change. However, the development industry is constrained by a system of planning policy, zoning negotiations, and administrative protocols that adjust slowly to persistent housing demand pressures. Migration adds workers and households, but this only matters if demographic growth is accompanied by increased demand for highly accessible and opportunistic locations. Building locations and credit determine cities' productive capacity. Cities also locate housing supply close to jobs, transport, and social networks, where agglomeration effects and competition are greatest. Migration can increase productivity by increasing output and wages and by increasing competition for space in cities. Supply responses to changing demand differ across urban centers. Even if there are no new arrivals, cities could respond to migration demand through rising prices, perpetuating the scarcity equilibrium. Even housing may simply become a store of value rather than an everyday way of gaining access, thereby reinforcing exclusion rather than being destabilized by new arrivals. This suggests that housing shortages should be considered less a function of migration-driven urban growth and more the collective effect of administrative decisions governing the growth process and its financial management. In this sense, migration-fueled narratives conceal regulatory and financial solutions through which cities are enabled to grow and develop.

Introduction

Immigration has been the focus of many contemporary political debates in Europe, especially when it comes to the issue of rising housing costs in major metropolises. Political discourse treats the increase in migration as an obvious and direct cause of a shortage of housing and presents it as a matter of demographics rather than a systemic issue. The durability of this account may be a function of the immediacy of demographic change in everyday urban life as opposed to the largely invisible institutional and financial infrastructure that actually shapes housing supply. It may also be due to the tendency for population change to function as both diagnosis and explanation, obscuring the processes that connect demand to prices, access, and exclusion.

The evidence across a wide body of economic literature suggests that immigration creates more income by increasing productivity, innovation, and labor market flexibility, and it increases

competition for space where jobs, services, and social opportunity are co-located. In general, because these pressures do not need to be driven by population growth, rising wages increase households' willingness to pay for access to constrained central locations, regardless of population growth. Immigration reshapes central housing markets, therefore, not simply by adding demand, but also by strengthening demand for privilege. The benefits of central locations, after all, are always limited and skewed.

Immigration is often singled out as the primary cause of housing shortage because its effects can be seen immediately and directly in increases in rents, more competition for rental properties, and changes to the make-up of the population. In contrast, capital allocation decisions, planning constraints and funding arrangements are mediated through structural legal and financial systems remote from ordinary experience. Where planning lags behind, where political forces block increases in density, and where preferential treatment for capital appreciation dampens the supply of housing, higher demand is not matched by more supply. Where this is the case, immigration is expected to be rare, not because we reach the limits of our capacities, but because our institutions do not absorb growth and maintain access.

European cities show a complex interaction between demand for housing that has been pushed by migration, investment practice and behavior, a city specific planning system and political interpretation. Immigration can lead to a rise in demand for housing in some European cities, but this can lead to either opportunity or exclusion depending on land, capital, and housing regimes. Where growth signals are largely capitalized into speculative investment, as opposed to demand-driven residential production, cities experience a housing affordability crisis even as economic growth is strong. Where planning systems incentivize certitude over adaptability, inelastic housing supply reflects endogenous or exogenous demand. When scarcity is attributed politically, at least to demographic change rather than institutional design, policies focus on migration and keep devices for exclusion in place. It is not just the size of the inflow but how it is institutionally framed that matters for the importance of immigration.

This paper intervenes in debates on housing shortages that ascribe the problems to demographic pressure by focusing on institutional mediation. While the effects of immigration on housing prices and neighborhood composition are well-studied, less attention has been paid to how institutional capacity influences the effects of immigration in European contexts. The synthesis provides an institutional framework that builds on these drivers, namely the demand structure, the capital available for housing investment, and the governing capacity of the city, to explain how immigration can produce housing shortages in some cities and not others, making affordability outcomes an institutional response to growth.

Contemporary European political discussions of housing markets in major urban areas are focused around immigration. Immigration-driven demographic growth is perceived as a natural cause of housing shortages and affordability problems rather than a feature of the housing market itself. In part, the persistence of this view is because demography manifests more directly in everyday life, in contrast with the largely hidden institutional and financial processes supporting housing supply. Population growth is used as both a point of diagnosis and explanation, sidelining the underlying processes of converting housing into prices and access and of its exclusion from some groups.

Many studies have found that immigration increases productivity growth, innovation, and labor market flexibility, which increase income and purchasing power. In turn, this may increase the demand for housing in urban areas where economic activity, infrastructure, and social opportunities are focused. This is important since pressures can exist even without population growth. If wages increase, households will be more willing to pay a higher price for access to limited central space. Immigration can influence housing markets not only by changing demand but also via increasing demand for unevenly distributed limited advantage.

Nonetheless, immigration is still cited as a cause of the housing crisis, with the immediate effects being rising rents, more competition for rental units, and demographic change among the existing population. Instead, the processes of allocating capital, planning constraints, and financial structuring are all mediated in the legal and financial systems, far removed from the daily lives of most people. When planning approvals are subject to bottlenecks, when densification is met with political resistance, and when capital markets are conditioned to value asset appreciation, housing supply is dampened. This leads to immigration as a scarcity condition not because migration swamps capacity but because institutions fail to absorb growth in ways that preserve access.

European housing outcomes are thus shaped through the interaction between migrant demand, the preferences of capital investors, the particular planning traditions of different countries, and the political framing of migration. The nature of the response of planning systems to increased demand is influenced by the regulation of land, the mobilization of capital, and the allocation of housing. In other cities, this growth signals speculative rather than incremental development, and the increase in supply is insufficient to stem rising housing costs as the economy continues to grow. In a closed planning system, supply is slow to respond to any demand. In populist discourses on scarcity, demographic change is the primary cause, not institutional arrangements. Politics then focuses on migration and its articulation while leaving the vehicles of exclusion untouched. Immigration, then, is not just about the influx of immigrants, but also how we interpret their immigration and respond to it.

This paper contributes to the housing supply constraint debate through an institutional mediation approach, building on studies that have shown effects of immigration on housing price and neighborhood composition in Europe. It also contributes to the literature on institutional variation. It proposes an institutional mediation framework that relates demand structure, capital flows, and governance capacity to derive an explanation of the differential effect of immigration on housing supply shortage across cities. In developing this conceptual and comparative institutional lens drawing on work in urban economics, housing finance and political economy, the paper argues that affordability outcomes are best understood as a response to growth rather than demographic pressure. It does so by examining how demand is distributed spatially, how capital understands population change and how the institutional capacity for housing supply varies across European cities.

Immigration, Demand, and the Spatial Structure of Urban Housing

In Europe, housing consumption is less a function of supply than of the concentration of opportunity offered by urban residences: employment agglomeration, transport infrastructure, educational institutions, and cultural amenities are bundled in urban areas in ways that cannot be easily replicated in peripheral areas. To summarize Wu's findings, households do not treat housing options as interchangeable across space. However, urban studies tend to view cities through the lens of spatial hierarchies in which location affects economic mobility and school outcomes. A spatial hierarchy can create strong competition for central urban areas, even when housing is available in the metropolis.

Rising income will also benefit this process by expanding the number of households that can access the limited amount of space available in the center. Immigration affects this process both through population increase and by providing more depth in the labor market and increasing wages. The demand is not only for greater numbers of units but also for demand with greater purchasing power for the same non-reproducible location. Wu notes that housing on the periphery cannot substitute for housing in the core without meaningful loss of utility, concluding that "peripheral housing rarely substitutes for central housing without significant loss of utility" (Wu). This explains why challenges of affordability persist in the core despite aggregate supply of the urban periphery. In fact, as neighborhood effects research shows, "children's outcomes in adulthood vary substantially depending on the neighborhood in which they grow up," which encourages people to cluster towards high-resource environments (Chetty, Hendren).

Scarcity arises not from a shortage of housing but from a shortage of access to high-utility locations. The cost advantage of lower prices at the urban edge rarely outweighs the loss of access in central locations. With increased income, households use their purchasing power to outbid for resources that provide longer-lasting economic and social benefits, rather than

increased square meterage. Immigration reinforces these trends as it sustains higher incomes. However, this effect is completely dependent on whether supply can keep up with demand in the cities and neighborhoods of choice. Regimes that constrain density or postpone development in central neighborhoods leave the displaced demand to push up prices in surrounding neighborhoods, including ones that have long been open to middle-class and working-class families. Housing remains, but affordable housing becomes scarcer as access to opportunity becomes more difficult to attain.

Demand structures likewise shape the political and economic interpretation of scarcity: if households have more money but still cannot afford housing near places of work or schooling, demographic change may seem like a plausible explanation. Demand matters more when housing supply is constant because the value of housing is not the quantity of housing, but its proximity to opportunity. While immigration contributes to urban wage and dynamism, it also pushes more of the regional household stock to compete for the same irreplaceable locations. Until we fully understand the structure of demand, immigration is likely to be a scapegoat. Demand, however, is defined by the availability of substitutes, not population growth. Hence, while demand conditions the nature of housing pressure, the way that housing pressure is translated into price, investment decisions and exclusions depends, in turn, on the way that capital responds to growth signals.

Capital Flows, Financialization, and the Transformation of Housing

Capital flows convert growing housing demand into return expectations. It has a stronger effect on urban housing markets than demographic changes. For instance, immigration signals productivity and income, and capital moves to take advantage of the benefits from residential real estate appreciation. Pavlov and Somerville argue that immigration “attracts outside investment capital”, which results in an increasing number of people buying housing due to asset value, rather than its use value. Prices can therefore adjust based on expected future population growth, rather than current population levels, transforming demand from a demographic into an economic phenomenon. Housing markets became arenas where households and investors competed for the scarce resource of housing to reap future returns.

Investment-driven development may expand aggregate supply, but seldom in accessible ways. In Europe, development finance has mostly focused on the project-level returns on investments, not the affordability of the projects; instead, as Squire finds, capital is directed toward “high-value projects that offer strong returns,” especially in markets where demand is forecasted to remain robust (Squires). Developers therefore concentrate on luxury units, high-rent apartments, and architecturally prestigious developments rather than housing accessible to lower-income individuals. While such projects increase unit counts, they do little to relieve scarcity at the bottom or middle of the market. New construction enters the housing stock as a

financial product, while existing units are priced upward through capital-driven competition. Immigration's influence becomes inseparable from this investment logic, as demographic growth is interpreted less as a planning challenge than as an opportunity for value extraction.

Capital flows exert structural power by reshaping how land and housing are valued, allocated, and withheld. Institutional investors have assumed a growing role in European residential markets, altering supply priorities and market behavior. As one analysis notes, "Institutional landlords accounted for nearly 40% of residential acquisitions in 2020, up from 10% in 2015", hence structurally improving financial capital's power over owner-occupiers (Dancygier). Viewing housing primarily as a store of value, investors are incentivized to hold units vacant, convert properties into short-term rentals, or delay transactions until prices peak. These strategies reduce effective supply even when construction activity remains strong. Affordability erodes because housing counted in physical units is not equivalent to housing accessible to residents.

The operation of capital markets remains largely invisible to households, who encounter scarcity through everyday struggles to secure housing. Demographic change is immediately observable, whereas investment decisions unfold through legal, financial, and institutional channels that few residents directly encounter. As a result, immigration becomes an intuitive explanation for rising rents even when speculative capital plays a more decisive role. This misalignment between perception and structure fuels political narratives that target newcomers rather than the financial mechanism converting demand into exclusion. Capital amplifies demographic pressure, but immigration becomes the symbol of scarcity because it is visible, while capital remains abstract.

This is also similar to cases across Europe, where financialization of housing means housing is no longer a site of shelter, but a global financial commodity. Financialization leads to forms of urban development where the production of value becomes more important than supply or, at times, even access to housing. In places where institutions to limit speculation do not exist or are weak, or where institutions to fund affordable housing are strong, demand for housing translates systematically into exclusion. When this is the case, financial flows should be a central explanatory lens for housing affordability in growing, prosperous cities.

Institutional Capacity and Governance Constraints

While demand for space by financial capital is unlimited, and is signaled in price, supply is constrained by the institutions governing land, housing and investment decisions and their underlying amassed social and political relations with the built environment created by capital. However, strong, unified governance can make this additional demand productive, while weak,

fragmented governance can lead to market signals accentuating exclusions for certain groups. Immigration does not, therefore, create housing pressure in the system; it provides a window into the limits of institutional arrangements managing growth. Transparency matters: when proper registration is clear, information asymmetries are reduced, speculation on opacity is lowered, and public scrutiny is allowed over land use and investment behavior. According to Walacik, “Transparent property registration reduces information asymmetry and stabilizes prices,” allowing governments to regulate transactions and enforce investors’ accountability more effectively (Walacik). Where transparency is an issue, property can be consolidated behind closed doors, ownership structure can be obscured, and housing can be treated as a financial product instead of a social good. The demand is met by accumulation for speculation, not for opening access; this is exacerbated by immigration, not by an overwhelming of the supply, but through the inability of institutionalized means to distribute the capital towards productive development. Planning regimes further shape housing system responses to demographic and economic change. Across Europe, planning regimes focus on heritage preservation, neighborhood stability, and gradual development, rather than rapid adjustment. These priorities preserved the form of urban areas, limiting density increases and land use transition to higher performing uses. Through empirical examples from Central and Eastern Europe, the role of institutions in Moatzed-Keivani argue that rapid privatization of state-socialist housing systems “produce uneven spatial outcomes that remain visible today,” capping the threshold of planning institutions to respond to later demand pressures (Ghanbari-Parsa and Moatazed-Keivani). In such contexts, immigration increases competition for high-utility locations, but scarcity emerges primarily from institutional rigidity rather than population growth itself.

Fiscal arrangements also complicate responses since the location of the benefits differs from the location of the cost. Immigration generates long-run economic rents at the national level, while costs of housing, infrastructure, and public services are incurred by lower-level governments in the form of municipalities. When the U.S. Congress states that immigrants “generate long-term economic benefits”, it implicitly recognizes the asymmetric time frames that local governments face in funding the demand on services that immigrants represent. When municipalities do not equitably share the revenue growth that comes with in fiscal burdens may reinforce fears that immigration is stretching capacity, when in fact its poor governance.

Institutional choices therefore shape material housing outcomes, as well as the frames through which scarcity is interpreted. When transparency and planning flexibility are limited, and fiscal authority is fragmented, residents perceive the material cause of future housing shortages to be a failure of social cohesion. Thus immigration becomes a proxy for institutional underperformance, even if the mechanisms of scarcity reproduction are obscured by the highly visible competition with newcomers, as well as the less visible regulations and tax regimes that translate demand into exclusion. Building institutional capacity is critical to a more equalizing

adaptability to growth, and the link between fiscal and economic gains, the political tension generated by immigration will arise not from a shortage of housing, but rather from failure to govern growth inclusively.

Public attitudes toward housing scarcity are conditioned not only by the material realities (scarcity), but also how individuals think about and explain the causes of scarcity. Locally, residents face higher rents and competition, and/or less occupation of space. For similar reasons, experiences encourage nearness and visibility based understandings of the causes of change, not necessarily the institutional and financial processes that govern housing markets. Immigration is the most visible of urban change processes, and hence the most normal, even if there are more applicable factors to exclusionary processes.

Political science research has found that economic uncertainty and demographic change increase support for exclusionary narratives. Haugsgjerd and Berd state that, “support for radical right parties increases when economic uncertainty coincides with demographic pressure” (Haugsgjerd and Bergh). Thus, housing scarcity can be a cultural signal; that is, housing scarcity can be a cultural and political rather than merely economic outcome. This is especially true for housing markets, where competition is local and visible (people see new arrivals to their neighborhoods), while other strains on the system (capital allocation, credit expansion and planning limits) are far away. Immigration is therefore the most obvious factor to blame for housing strains, even if most of the causes are systemic.

While those public pressures are compounded, the financial arrangements behind them are more obscure, drawing attention to scarcity without providing an immediately legible explanation: credit and speculative finance increase price fluctuations through less than quotidian channels. As Zurek notes, (increases credit provision on the assumption that prices would continue to rise indefinitely,” accelerating price growth even in the absence of proportional population increases (Zurek), but these mechanisms rarely enter public discourse. Instead of the financialization of housing, the shortage of housing is explained through narratives of demographic completion, which prove intuitive explanations for individual dependence.

Resistance to development reinforces these interpretive patterns. Opposing density expansion based on neighborhood character, congestion, and social change limits supply in the places of greatest demand. However, when affordability does decline under these conditions, it is due to immigration because this coincides with population growth and opposition to access. Yet opposition to developing new housing began before recent immigration, and persists regardless of demographic change, thus limiting supply in its own right. Framing the issue of scarcity in terms of population pressure instead of institutional un-development shifts the political focus away from exclusion.

These dynamics influence policy response: when housing scarcity is framed like a demographic problem, states may strengthen migration regimes while leaving other levers (like financial incentives, planning regimes, and institutional barriers) unchanged, and respond to a visible symptom rather than the structural factors underlying exclusionary outcomes. Immigration is a symbolic fix to a market design and governance problem. Unless the scarcity of affordable housing is a problem of institutions, immigration will exacerbate social conflict but have little impact on affordability.

Political interpretation therefore occupies that last link in the causal chain between the structure of demand, capital flows, institutional capacity, and housing outcomes. Scarcity is politically important not because it leaves housing inaccessible, but because the processes that produce exclusion do not. Immigration, being present, proximate, and manifest, works as a

scapegoat. The relevant capital and institutional design are different, less immediate structures of feeling. A theory of housing affordability must be both structural and interpretive and shift the political gaze away from the immigrant demographic and toward the political economic structures that rent urban opportunity.

Institutional Mediation Across European Cities: Historical Evidence

Over the last century immigration has shaped European housing markets. Experience shows that the affordability impact of immigration is less a function of the scale of demographic pressure than of how national and local governments organize and channel this pressure via planning, housing finance, and public provision. But in periods of high migration, divergent housing outcomes are observed across European cities, suggesting that migration is a multiplier of institutional arrangements.

The peak in industrialization in the late nineteenth and early twentieth centuries also resulted in a large influx of rural and overseas workers to some major European cities, whose populations increased rapidly: in London, Brussels, and Milan, the effect on housing was quite different. London's neglect of social housing and reliance on markets created spatial polarization. Brussels experimented with municipal housing, which did alleviate exclusion in working-class neighborhoods, but Milan allowed private developers to dominate, resulting in a highly unequal spatial structure. In all these cases, hiring immigrants raised demand, but economic shortages were most severe when public intervention and organized planning failed to keep pace.

Similar events occurred during 20th century reconstruction. A sudden influx of people does not always result in housing shortages when the framework for institutional capacity is present. Labor migration from Southern Europe, North Africa, and Turkey reshaped cities like

Stuttgart, Geneva, and Rotterdam after the Second World War. Despite the tight competition for space, the long-term scarcity was offset by other institutional factors. The examples of Stuttgart and Geneva include housing for workers as part of the reconstruction and infrastructure. Meanwhile, in Rotterdam, abundant municipal housing absorbed the migrant influx, and although demand increased, affordability was maintained as it was seen as a key recovery by governments.

European cities in the late twentieth century were less involved in the direct provision of housing. The net migration into urban housing markets of Lisbon, Dublin, and Barcelona, all of which took place in the late twentieth century, occurred under institutional conditions distinguished from the preceding decades. In Lisbon, deregulation enabled speculation that favored wealthier households over lower-income (including recent immigrants). In Dublin, the limited public housing stock permitted speculation that further raised prices. In Barcelona, increased resilient development countered dominant speculation, but tourism and renewed migration to the city remained a pressure on central neighborhoods. Tourism and migration strengthened affordability pressures when public capacity receded and market allocation had greater weight than public policy.

These institutional policies also apply to more recent massive waves of migration, starting in the mid-2010s. Berlin experiences fast rent inflation, together with more refugees and labor migrants. However, the country's housing shortage largely results from long-term structural factors, like social housing privatization, deregulation, and slow permitting processes, rather than immigration. By contrast, Vienna saw its sustained immigration offset by its continued expansion of municipal and limited-profit housing to make growth affordable. In the mixed cases (Barcelona, Dublin), migration has interacted with tourism, capital inflows, and restrained public supply to move what is remaining much further into the domain of more attractive areas. In these cases, migration is not the driver of the housing shortage, but rather a multiplier of the supply, access, and land use institutions already in place.

For example, in Central and Eastern Europe, a legacy of rapid privatization in the 1990s shaped the spatial structure of post-enlargement migration from the 2004 and 2007 European Union accessions. Ghanbari-Parsa and Moatazed-Keivanai argue that privatization created “uneven spatial outcomes that remain visible today,” for the regimes of migration absorption in the cities of Prague and Budapest. Rather, systems lacking in social provision and regulatory capacity struggle to extend access as mobility opportunities increase. This trend towards competition largely within the most desirable districts was due not simply to immigration creating scarcity, but rather the institutional infrastructure to redirect growth towards access being eroded long before population pressure.

The third type of institutional mediation is represented by the European Union's free movement policies; migration from Poland, Romania, and the Baltic states accelerated urbanization in Germany, the United Kingdom, and Northern Europe. These countries had very different housing outcomes, with Vienna, for example, expanding its municipal housing system to integrate across incomes. Copenhagen's planning chose higher density to accommodate growth, reduce land value speculation, and maintain access to opportunity, whereas Hamburg and Munich's strategies involved lower-density growth and market allocation, leading to larger losses in affordability with increased demand. Taken together, these findings underscore that mobility increases demand for housing and that its affordability depends on effecting development-oriented planning systems.

A survey of European experiences pointed to a number of lessons. Immigration magnifies both the strengths and weaknesses of institutions. Where cities expand supply, regulate speculation, and invest in public provision, immigration can support economic dynamism and social cohesion. Without these measures, cities risk becoming scarcer, poorer, and more divided. Where development is restricted, and housing is only built if the private sector deems it feasible, any additional demand, whether from abroad or from internal migration, will be exclusionary. The last 100 years of European housing have shown how institutions shape if a growing population is a source of opportunity, or a source of conflict. Immigration identifies such conditions, but does not dictate their outcomes.

While European cross-border migration complements local labor, wages and productivity, and long-run growth, its housing outcomes depend on the interaction with demand structure, capital mobility and institutions. An increase in income entails more bidding households for the same high-utility housing in neighborhoods with great employment, education, and transportation productivity growth with expectations of asset price appreciation, accelerating the rise in house prices and construction above actual housing needs. Whether such a response represents an opportunity or a constraint on housing supply depends not on the type and amount of demand (or population growth), but rather on planning and fiscal and regulatory frameworks. Immigration matters, therefore, not because it overwhelms housing markets, but because it enables the institutions to mediate the transformation of growth into access or scarcity.

Restoring affordability by restricting immigration misdiagnoses the structural exclusionary effects on high housing prices. Speculation, limited land supply, and fragmented governance will continue to constrain access to high-opportunity neighborhoods, even in the absence of demand-led price growth. The European experience shows that housing shortages are most likely where institutions retreat from direct provision, allocate access to capital, or restrict density in those places where demand exceeds supply. Immigration is only associated with housing shortages when these institutional arrangements are unable to provide access to housing

in a context of rising prosperity. Yet while institutional responses stagnant, demand concentrates, capital increases competition, and exclusions are enacted regardless of demographic trends, targeting migration merely addresses a visible symptom of scarcities rather than their structural causes.

Setting the conditions for growth that can expand affordable housing opportunities for all requires governments that are willing and able to govern inclusively. Planning, taxation, and regulation must be coordinated, to increase density where opportunity is concentrated, assign costs where they will be borne, and ensure that investments expand access rather than extract value. The same applies to politics, which needs to get beyond demographic reductionism to get to grips with how housing markets operate in a growth environment, and what immigration means in relation to these institutions. Where those frameworks are weak, immigration magnifies exclusion and local social tensions. Where they are strong, immigration reinforces productivity, mobility and social cohesion. Housing justice in Europe has not been about controlling flows, but whether cities are prepared to find ways to translate economic energy into equitable access to opportunity that lasts.

Works Cited

Accetturo, Antonio, Massimo Bugamelli, and Andrea Lamorgese. "Don't Stand So Close to Me: The Urban Impact of Immigration." *Regional Science and Urban Economics*, vol. 45, 2014, pp. 45–56. Accessed 3 Dec. 2025.

Ghanbari-Parsa, A. R., and R. Motazed-Keyvani, Braakmann, Nils. "Immigration and the Property Market: Evidence from England and Wales." *Real Estate Economics*, vol. 47, no. 2, 2016, pp. 509–533. Accessed 7 Dec. 2025.

"Development of Real Estate Markets in Central Europe: The Case of Prague, Warsaw, and Budapest." *Environment and Planning A*, vol. 31, no. 8, 1999, pp. 1383–1399. Accessed 11 Dec. 2025.

Haugsgjerd, Atle, and Johannes Bergh. "Explaining the Surge of the Populist Radical Right: A Time-Series Analysis of the Effects of Immigration and the Economy in Norway." *European Political Science Review*, vol. 16, 2024, pp. 209–225. Accessed 14 Dec. 2025.

Pavlov, Andrey and Tsur Somerville. "Immigration, Capital Flows, and Housing Prices." *Real Estate Economics*, vol. 48, no. 3, 2020, pp. 915–949. Accessed 18 Dec. 2025.

Squires, Graham, Neil Hutchison, and Michael White. "Innovative Real Estate Development Finance: Evidence from Europe." *Journal of Financial Management of Property and Construction*, vol. 21, no. 1, 2016, pp. 54–72. Accessed 22 Dec. 2025.

Walacik, Marek. "Property Appraisal via the Lens of Property Registration Abundance: Real Estate Market Asymmetry Assessment." *International Journal of Strategic Property Management*, vol. 28, no. 2, 2024, pp. 390–404. Accessed 26 Dec. 2025.

Zurek, Maximilian. "Real Estate Markets and Lending: Does Local Growth Fuel Risk?" *Journal of Financial Services Research*, vol. 62, no. 1, 2022, pp. 27–59. Accessed 30 Dec. 2025.

How Does Exposure to War Affect Children's Physical and Mental Health, and What Strategies can be Implemented to Protect Them? By Viktoria H. Gilbert

Abstract

The intensity and frequency of the armed conflicts during the first quarter of the 21st century have sent civilian deaths surging 40% globally in 2024. According to the United Nations Annual Secretary General Report on Children and Armed Conflicts, violence against children in armed conflict in 2023 reached a shocking 21% increase, including especially grave violations. War-affected children develop a wide array of psychological and physical health problems that, without timely and structured interventions, can persist throughout the duration of the child's life. The scope of this review is to provide a comprehensive and current literature review of war effects and supportive interventions for war-affected children. To supplement this paper, systematic research in PubMed, Medline, Google Scholar, Google Books, and online open libraries of the United Nations, United Nations Children Fund, and Safe Children has been conducted.

The study included 101 primary and secondary studies published in English between 2000 and 2025, addressing both direct and indirect war effects on children, their families, and communities. Additionally, various intervention strategies, as well as international organizations and the world leadership approach to addressing war-inflicted adversities on children, have been included. Research findings indicate that while many studies have been done on war effects on children, a minimal number of studies highlight the importance of structured and overarching approaches to interventions addressing the more complex psychological effects of war. With the continued escalation in intensity of conflict worldwide and the increase in government military expenditure, it is imperative to continue documenting and studying the physical and mental health effects of war on children, and most importantly, to continue to identify and advocate for strategies that can be implemented to protect war-affected children and their rights.

Introduction

The 20th century had the greatest number of recorded wars and was the most lethal for civilians. Since the beginning of the 20th century, the civilian population has been repeatedly targeted during wars as one of the strategies of political struggle (Lynch, 2012).

During 2023, 59 armed conflicts were recorded, peaking in 2024 with 61 armed conflicts, most of which were internationalized civil wars, in which other countries participated by providing political, military, and/or financial support (proxy wars) (High Commissioner for Refugees"). While the wars in Ukraine and Gaza have received much media attention, most armed conflicts are "out of sight and out of mind" for many people in Europe and the United States (Levy, 2025).

New data released by the UN Human Rights Office shows that civilian deaths in conflicts surged 40% globally in 2024. Every 12 minutes, a civilian dies in armed conflict today (UN

Human Rights Court, 2025). Overall, it is estimated that there have been 21.8 million war-related civilian deaths since WWII; the Red Cross estimates a number twice as high, approximately 40 million people (Lynch, M. F. 2012). The Office of the United Nations Human Rights Court recorded that 21,480 women and 16,690 children were killed in conflicts during the 2023-2024 period, with 80% of child deaths and 70% of women's deaths recorded in Gaza alone, which is an increase of approximately four times more children and women than were killed in armed conflicts when compared to 2021-2022. Overall, violence against children and women in armed conflicts has had a devastating upward trend over the last two years (2023 and 2024) (UN Human Rights Court, 2025).

According to the United Nations annual Secretary General report on Children and Armed conflicts, violence against children in armed conflict reached extreme levels in 2023, with a shocking 21% increase in particularly grave violations (General Assembly Security Council, 2024). Peace Research Institute Oslo (PRIO) reported in December 2024 that over 473 million children (one out of six children globally) now live in areas affected by war or conflict (PRIO, 2024).

The United Nations definition of war is identified as “a formalized and declared armed conflict between armed forces of states or governmental forces or groups” (“A New Era of Conflict and Violence”), vs. a armed conflict, as described by Red Cross, is generally defined as a situation of organized collective violence that affects relationships between society or power relations within society, a state of hostilities dependent on neither a declaration nor recognition of the existence of “war” by its parties (“How is the Term ‘Armed Conflict’ Defined in International Humanitarian Law?”). The effects of war are catastrophic for the combatant countries and human development (Vesco et al., 2025). The following sections review existing literature on how war-related experiences affect child development in this context.

Despite many repeated calls from UN General Assembly Security Council, United Nations International Children’s Emergency Fund (UNICEF), World Health Organization (WHO) and other involved parties to implement measures to end grave violations against children in the Gaza, Israel, Afghanistan, Central Africa and other conflict zones, the dramatic increase in reported cases of violence against children remains terrifyingly high (General Assembly Security Council. 2024). These facts and staggering statistics really bring to light the devastating toll that armed conflicts have on vulnerable child populations. It also emphasizes that there is an overall urgent need for stronger protection measures for civilians against war, the need for prompt accountability, research, and advocacy for peace.

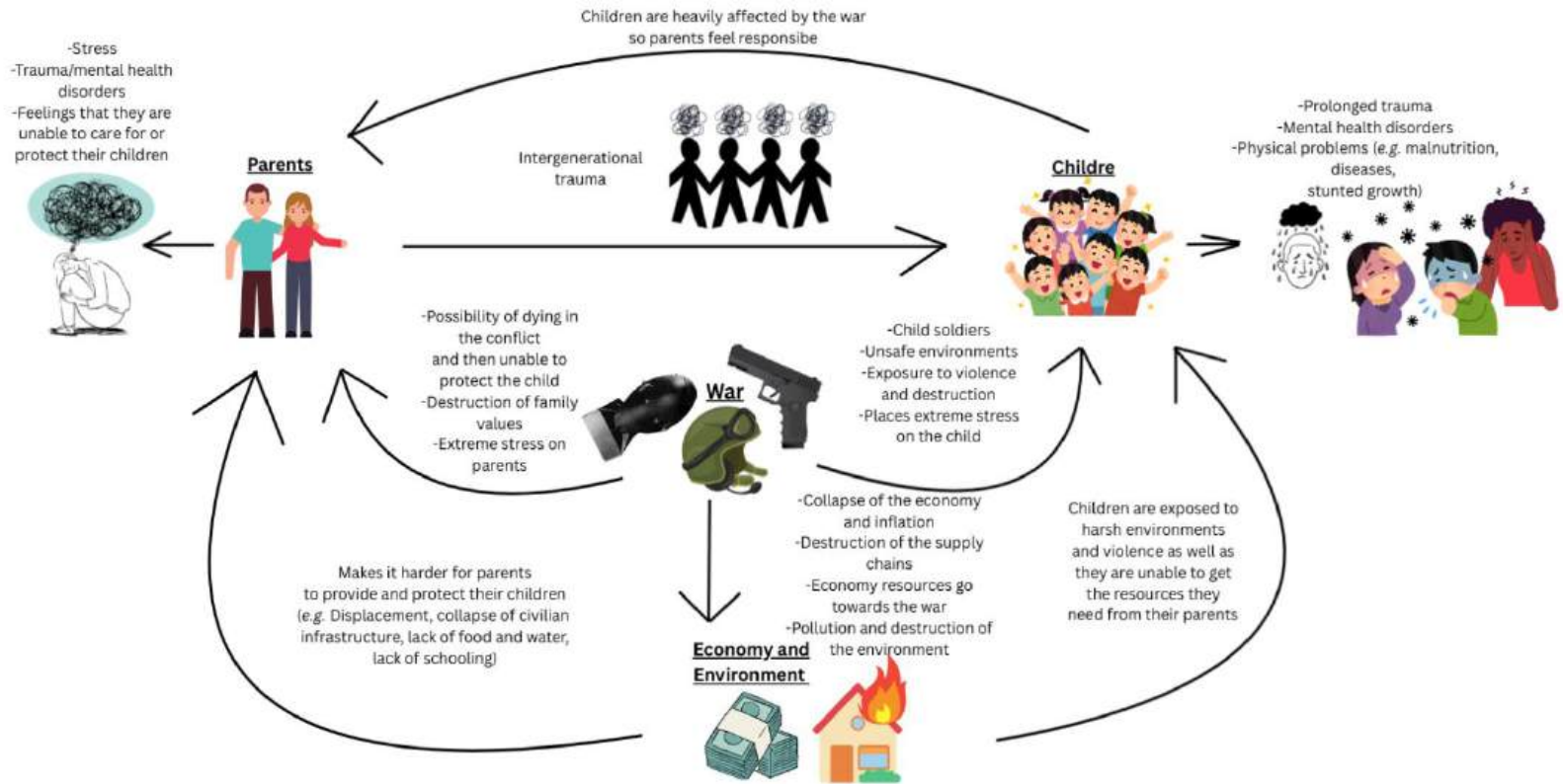


Figure 1.1. Depiction of the many effects and impacts war has and how it creates a “ripple effect.”

Studying children for war-inflicted traumas is important for future health policy and humanitarian responses

In the context of this study’s analysis, a child is an individual under 18 (or 19) years old in a period of rapid growth and development from birth to adulthood, encompassing infancy, childhood, and adolescence (Committee on Improving the Health and Well-being of Children and Youth through Health Care System Transformation et al., 2024). Children are the most vulnerable group during armed conflicts, and the effects of war are catastrophic and long-lasting for their development and overall well-being (Peltonen, 2025). With the highest number of conflicts recorded since WWII (Global Peace Index, 2024), it is estimated that more than 19% of children in the world are currently living in a war zone. Furthermore, almost 50 million children have been forcefully displaced due to conflict and violence (UNICEF, 2025).

Children often face psychological hardships during armed conflicts. These hardships include forced family separation or loss of family members, loss of access to education and healthcare, insecure access to food and shelter, and forceful displacement (Betancourt et al., 2015). Additional factors include having their rights violated, various acts of violence and discrimination, danger of being killed or injured, vulnerability to disease, and not receiving life-

saving vaccines, critical malnourishment, and more (Martini et al., 2024; Betancourt et al., 2015). All of these ordeals establish the premise that civilians affected by war or armed conflict do not only experience physical trauma; In particular, children go through many psychological traumas that could have long-lasting effects on their livelihood and the lives of their families. This includes Post Traumatic Stress Disorder (PTSD), anxiety, depression, somatic stress disorder, and other conditions. Studies have reported that among children living in war zones, child refugees and asylum seekers or children who were forced to leave their home country because of a serious threat to their life but were not promptly granted refugee status (“Asylum-seekers”), have a 22.7% prevalence for post-traumatic stress disorders, 13.8% for depression and 15.8% for anxiety disorders within these populations (Peltonen, 2025). There is also an increasing awareness that adverse conditions may continue through intergenerational biological and psychological mechanisms in war victims, and this could spread the effects across generations and show in their children as well (Negriff, 2020). To date, large-scale recovery efforts have been focusing on addressing visible and immediate needs of civilian war victims (food, water, and shelter), psychological health, and economic aid. Psychological trauma is usually not addressed promptly, which leads to these issues becoming part of society and future generations.

Physical and mental health is a precious asset, essential for both individuals and communities (Martini et al., 2024). Maintaining the good health of children is imperative for future development and society (Clark et al., 2020). When countries sacrifice the health of their children, they tarnish their own future progress potential for generations to come. According to the UN, military spending worldwide reached a staggering \$2.7 trillion in 2024 and intensified political tensions; there is a critical and urgent need to protect millions of children in conflict zones, and the world’s organizations and governments must take this situation very seriously, honor their commitments, and reduce the impact of war on the most vulnerable population (UN, Andersen C., 2024). Humanitarian world organizations, governments, and world leaders can employ several strategies on both global and more localized levels to improve protection for children from war. Global response recommendations are: prioritizing long-term solutions for peace and conflict prevention, uphold standards of conduct in conflict, hold perpetrators accountable, enforce credible sanctions on armed groups to curb violations against children, prioritize and advocate for humanitarian actions while avoiding politicization, support and cultivate children’s resilience and recovery efforts and listening to and meaningfully involve children into policymaking process related to children.

Localized recommendations for mitigating the humanitarian crisis issue are: prioritize children’s needs early in conflict and peace negotiations, scale up age appropriate humanitarian aid for children and their families (vaccines, child friendly nutritious foods and hygiene items), scaling up sustainable and multi-layered support for children and their families’ physiological and mental health, establish safe spaces for children and their families in conflict areas, increase

aid to rebuild critical civilian infrastructure (education and healthcare facilities), and promote systematic research on how armed conflict impacts children. Those strategies and more have been discussed by world leaders and the UN (Children and armed conflict. Report of the UN Secretary-General, 2024). Statistics show that there is still a lack of commitment and implementation to protect children.

Thesis statement: It is imperative to continue documenting and studying the effects of war on children's physical and mental health, while continuing to identify and advocate for strategies that can be implemented to protect them.

Background

Overview of child development and health needs.

Child health is a multidimensional status that also changes depending on the age of the child and their individual needs. The healthy development of a child requires nurturing and a stimulating environment that meets their essential needs, such as nutrition, clean water, safety, housing, healthcare, education, family environment, social and structural services support, as well as the psychological, cultural, social/communal, spiritual, and intellectual resources. Children should also be afforded the opportunity to have time to mature into adulthood (Szilagyi & Schor, 1998). The United Nations Convention on the Rights of the Child Article 6 (Article 6, UNICEF) states that every child has “a right to live...and develop healthy” and that every child has “the right to a standard of living that is good enough to meet their physical and mental needs” (Article 27, UNICEF). Children's health and well-being are greatly dependent on the care they receive from their families and communities. “States Parties shall ensure to the maximum extent possible the survival and development of the child,” as capable children are the foundation of a prosperous and sustainable society (Article 6, UNICEF).

Effects of war, violence, and prolonged stress on children's development.

People go through several sensitive developmental periods during their childhood and adolescence, which exert a great influence on brain development and behavior (Siehl et al., 2022). During these sensitive periods, traumatic experiences and stress can cause inadequate developmental outcomes, as early experiences of stress may be consolidated over time (Murphy et al., 2022). Both processes, temporary and one in which stressors and trauma accumulate over time, negatively impact a child's physical and psychological development, and the consequences of experienced trauma and its effects are likely to transfer far into adulthood even after the conflict itself stops (Samara et al., 2020).

Children impacted by war experience a large array of trauma. The ways in which children are impacted by war can be divided into two interconnected categories. Direct war-related experiences, such as shelling and bombing, violence, and forced evacuation, displacement, injuries, malnutrition, etc. (Gredebäck et al., 2022). Indirect war-related traumas are associated

with parents, family, teachers, social networks, and other significant adults or relations that were impacted by war trauma and prolonged stress, which may have direct negative effects on the child's development (Samara et al., 2020).

Several studies indicate that children who have experienced war display elevated levels of psychopathological and psychosomatic development problems compared to those of non-war-exposed children (Kandemir et al., 2018). Studies report that among war-affected refugee and asylum seeker children, there is 22.7% prevalence of Post-Traumatic Stress Disorder (PTSD), 13.8% for depression, and 15.8% for anxiety disorders in these populations (Blackmore et al., 2020). A recent study of Syrian refugee children and adolescents resettled in Lebanon and Jordan reported a high prevalence of PTSD and an elevated risk of emotion dysregulation (e.g. significant mood swings, feeling of panic, anger, sadness and fear), with PTSD symptoms and the inability to control or regulate emotional response associated with social withdrawal, self-criticism, and feeling of resignation (Khamis, 2019). Chronic stress can have profound neurobiological effects, negatively affecting brain development and potentially leading to difficulties with attention, learning, and emotional regulation later in life (Qouta et al., 2021). Moreover, prolonged stress can weaken immune system function, disturb sleep patterns, hinder somatic growth, and elevate the risk of both acute health problems and chronic diseases (Ferrara et al., 2025).

Indirect war-related psychological trauma in children is often associated with their parental mental health (Eltanamly et al., 2021). Families are impacted by experiencing trauma and by having to live under a lot of daily stress. Such difficulties could change parenting practices and routine, which negatively impacts parental stress levels and mental health. Parents with mental health challenges may become irritable and distant toward their children; when parents need to constantly focus on keeping themselves and their children safe, they may have less time and mental space to tend to their children's emotional needs (Hazer & Gredebäck, 2023). Results suggest that the adverse effects of past war trauma and ongoing stress of war on parental general mental health, especially a mother's, can increase the risk of negative parenting behavior, and in turn contribute to poorer psychosocial outcomes for children such as worsening of post-traumatic stress symptoms, depression and anxiety, social problems, externalizing behavior, and child's post war adjustment (Eltanamly et al., 2021; Mattelin et al., 2024). For example, children develop their coping styles in the context of observing and interacting with their parents. Therefore, parent's coping styles determine their children's responses and influence their coping strategies (Kiser et al., 2008). The result is that the parents' psychopathology is directly reflected in the child's psychosocial adversities, even many years after accounting for children's war trauma exposure (Castro-Vale et al., 2019).

War-inflicted traumas and stress, such as coping with the feelings of fear and continued threat, as well as engaging in activities aimed at physical survival, can be thought of as additional

developmental tasks that take up resources that would otherwise be used for meeting age-typical developmental milestones during sensitive developmental periods of childhood and adolescence. These cumulative stressors are associated with poor mental health and developmental delays in several aspects, including cognitive functioning, emotion regulation, affective processing, and prospective control. Moreover, some studies suggest that the lack of safety and early life stress changes the neurological and immune systems, which enable motivated psychological and behavioral responses to the environment. The altered activity of these systems is associated with adverse mental and physical health outcomes following stress exposure (Peltonen, 2025).

These cumulative stressors are associated with poor mental health and developmental delays in several domains, including cognitive functioning, emotion regulation, affective processing, and prospective control (Hazer & Gredebäck, 2023).

Historical examples: Syria, Palestine, Ukraine.

One in four children live in a country affected by conflict or disaster (UNICEF, 2019). Generally, war-affected children face adversities such as forced family separation, loss of access to school and healthcare, insecure access to food and shelter, and displacement from homes and communities (Betancourt et al., 2018). Some of the aforementioned countries in this study have a long history of political, economic, and military instability that extended well beyond the war zones, facing a generational catastrophe, affecting nearly every facet of people's lives, but most importantly, profoundly affecting the lives of millions of children (Betancourt et al., 2020).

(i) *Syrian Civil War* (March 2011 – present)

The Syrian civil war is an armed conflict that began with the Syrian revolution in March 2011. There was large-scale general dissatisfaction with the Ba'athist regime ruled by Bashar al-Assad. This provoked many large-scale protests across Syria. As a result, the Assad regime retaliated with lethal force. This led to a series of uprisings by armed groups and civilian retaliation and ultimately, a civil war. The war lasted almost 14 years, but discontent and war effects prevail. The Assad regime and Syrian civil war are a grave example of the devastating effects on the civilian population, most notably on vulnerable children who witness acts of violence and war atrocities daily (Yayan, 2018). It is estimated that 2.4 million Syrian children were forcibly displaced due to the ongoing war in Syria ("Syria Emergency"). Millions of children fled to neighboring countries such as Jordan, Turkey, Egypt, and Iraq ("Syria Crisis Fast Facts"). War-affected Syrian children suffer from significantly high rates of mental illnesses, among the most frequent of which are depression, anxiety, and PTSD (Rizkalla et al., 2020). In Lebanon and Jordan, 45.6% of Syrian refugee children and adolescents developed PTSD, while 53% of Syrian refugee adolescents in Jordan experience emotion dysregulation and high insecurity levels (Khamis, 2019). It is also stated that 82.5% of refugee children and adolescents were exposed to more than four lifetime traumas (Panter-Brick et al., 2018). Syrian refugee children also experience emotional and behavioral regression, including high levels of anxiety,

bedwetting, nightmares, and panic attacks (Khamis, 2019). The prevalence of PTSD is estimated to be 10 times higher among Syrian refugee children than in children in general across the globe. At the same time, Syrian refugee adults suffer from high rates of physical and mental health sequelae, including PTSD and somatic symptoms due to war trauma and displacement hardships. In summary, both Syrian refugee children and adults have an elevated risk of psychopathology compared to the general world population. It is hypothesized that Syrians, both adults and children, have higher prevalence rates of mental health issues when compared to other traumatized groups due to a very large array of different traumatic experiences endured for an extended period of time (Hazer & Gredebäck, 2023). The need for mental healthcare within Syrian refugee children and their families is of great concern, especially given that armed conflict in Syria has been going on for 15 years, and an end to the conflict is not likely within the foreseeable future.

(ii) *The Israeli-Palestinian Conflict (1947 – present)*

The existing Israeli-Palestinian conflict is an ongoing military and political conflict over an initiative to seize the land within the territory of the former Mandatory Palestine. The conflict's origins rise from Zionism in the late 19th century in Europe, a movement that was driven by establishing a Jewish state through the colonization of Palestine. Children living in war zones are at high risk of developing PTSD and other emotional disorders, and children who are going through the Palestinian crisis are no different. Approximately 59% of Palestinian children who have been directly affected by war, i.e. homes had been bombarded and demolished during the conflict, reported symptoms of post-traumatic stress. Palestinian children who have had secondary exposure to war, mostly through parents and other important adults in their lives, reported high levels of anxiety and cognitive expressions of distress. In Palestinian children, (41%) reported moderate to severe post-traumatic stress reactions and high rates of anxiety and behavioral problems (27%). Studies have established a high frequency of problems such as fears of leaving home (28%), fears of soldiers (47%), and nightmares (7%), during the same period of political and military violence. The most frequently reported items of the Complex Post-Traumatic Stress Disorder - Reaction Index (CPTSD-RI) in exposed children were: identified event as extremely stressful (66%), difficulty with concentration (58%), sleep disturbance (57%), and avoidance of reminders (52%). The severity of emotional problems in Palestinian children living in the midst of fighting is evident upon examination (Thabet et al., 2002). Despite continued efforts at brokering peace, conflict has persisted ([Israeli-Palestinian Conflict Timeline | CFR Education](#)).

(iii) *Russo-Ukrainian War (February 2022 – present)*

The roots of the Russo-Ukrainian War began in February 2014 when Russia invaded Crimea, and eventually annexed the region. Over time, small disputes and armed conflicts ensued in Eastern Ukraine, but on February 24, 2022, Russia launched a full-scale invasion of Ukraine. According to a UN Human Rights Office Report, death, injury, and lasting family

separation are among the traumatic events that have upended the lives of millions of Ukraine's children in the three years since the Russian Federation launched its full-scale war. As of December 2024, it is estimated as of December 2024 that 737 thousand children had been internally displaced due to war violence, and 1.7 million children have become refugees, many of them separated from a parent, typically their father ("The Impact of the Armed Conflict and Occupation on Children's Rights in Ukraine"). As of June 2023, 35% of Ukrainian parents report that their children display higher levels of fear and worry, over 25% reporting higher levels of sadness, and 25% of parents stated that their children display more problems with concentration (Elvevåg & DeLisi, 2022). Studies of the adult population of Ukrainians have determined that 15% of adults met criteria for a diagnosis of PTSD and 26% met the criteria for a diagnosis of CPTSD (Karatzias et al., 2023), along with elevated levels of symptoms of anxiety, depression, loneliness, as well as alcohol and substance abuse since the beginning of the war. As stated by the first lady of Ukraine, Olena Zelensky, "[a]s research shows, unfortunately, 44% of children in Ukraine show signs of potential PTSD" ("Future Index").

Taking into consideration that the 2022 Russo-Ukrainian War is relatively new, researchers made the assumption that mental health problems will increase within the general Ukrainian population as the war is prolonged, especially among children and adolescents (Elvevåg & DeLisi, 2022). UNICEF anticipates that over 2.2 million Ukrainian children are affected by this war and need access to mental health and psychosocial support ("Future Index").

Based on the historical data provided, it is safe to say that exposure to the traumatic events of war, post-migration challenges, and other psychological stressors is the cause of increased risk of mental health disorders, such as depression, anxiety, and PTSD, in conflict-affected populations across the globe (Fazel et al., 2005). PTSD is among the most prevalently reported and extensively studied psychological outcomes in children affected by armed conflict (Bendavid et al., 2021).

War and conflict effects on healthcare, the education system and other civilian infrastructure, as reflected on children and the general population.

Destruction of educational infrastructure and schools during war often causes diminished literacy, poor academic performance, loss of educational potential, and academic achievements within children (Bharati, 2022). According to UNICEF, it is estimated that more than 52 million children in countries affected by conflict are out of school. For example, children in Gaza have missed out on more than a year of school due to armed conflict ("Not the new normal"). Exposure to conflict decreases the probability of passing final academic year examinations and university admittance (Brück et al., 2019). Lower educational levels in the younger generation contribute to war-induced losses in human capital, with long-term impacts on the productive capacity of an economy. Some studies have indicated that low education is a risk factor

associated with psychiatric disorders, low self-esteem, and feelings of hopelessness for the future (Gredebäck et al., 2022).

Healthcare systems' disruption during war has immediate and often irreversible effects on people's health and lives, resulting in excessive deaths, injuries, and disabilities. The deterioration of health systems during armed conflict is characterized by the destruction of physical infrastructure, disruptions in supply chains, and the diversion of state funds from health to the military. The lack of healthcare facilities and medical services often leads to disrupted treatments and delayed diagnoses of chronic and non-communicable diseases, pathogen outbreaks, and infection spread. These things are aggravated by the limited access to life-saving vaccines and medications within the war-affected population. Many studies noted that the indirect war deaths often surpass direct war death tolls because of the lack of healthcare, with indirect mortality accounting for at least 75% of the total excess deaths in armed conflicts (Bendavid et al., 2021). Due to loss of healthcare infrastructure, lack of qualified personnel, equipment, and medicine, restricting access to health services for mother and newborn increases mortality in the civil population UNICEF 8. For example, 10% of all conflict-attributable deaths among women are due to maternal mortality, likely resulting from deteriorated health infrastructure, which could be avoided in a time of peace (Vesco et al., 2025). Previously eradicated, vaccine-preventable diseases may reemerge in conflict-affected areas. Lack of immunization during conflicts and war, mass movements, and communal life in camps, refugee children are more susceptible to death from infectious diseases, especially when accompanied by malnutrition (Çelik & Özpınar, 2017). Over the past several decades, healthcare facilities and health workers have become direct targets during war and conflict, which directly impact people's ability to access essential health services, increasing the impact of war on children ("Stopping attacks on healthcare").

Modern military warfare also disrupts life-supporting ecosystems and inflicts degradation upon natural environments and biodiversity, which can lead to humanitarian crises, exacerbating societal tensions and conflicts, further complicating post-conflict recovery efforts (Kotsis, 2024). Disruption of the natural environment carries significant threats to civilians and public health. Due to pollution of water supplies in Gaza, up to 70% of its population reported consuming salinized or contaminated water, which poses a severe risk of dehydration and the spread of infectious diseases (Abuawad et al., 2025).

War has direct and indirect environmental impacts, ranging from the widespread destruction of vulnerable ecosystems and wildlife loss to waste generation and lasting pollution of vital natural resources. Direct impacts, like deforestation and soil erosion, are the most cited environmental impacts of war (Meaza et al., 2024). Moreover, indirect impacts like mental and psychological damage, cultural cost, and loss of identity that represent a connection to the land are rarely mentioned. One such example is the destruction of the Palestinian environment and the

uprooting of wildlife, olive tree orchards, and farms that have devastated many Palestinians, as it represented their deep connection to the land and cultural heritage (Abuawad et al., 2025). At the same time, the destruction resulting from military operations in Ukraine not only devastated local natural habitats and ecosystems but also undermined vital economic resources of the local population that rely on those ecosystems. Food security of impacted local communities and their livelihoods was disrupted when their fertile land was turned into military war zones and mine fields. It reduced the ability of the Ukrainian agricultural sector to contribute to the stability of the world's food security overall, causing food shortages throughout the developing world (Leal Filho et al., 2023).

Hardships and the impact of war go beyond physical destruction, affecting socio-psychological outcomes such as social capital, cooperative behaviors, and pro-sociality within the affected population. The impacts of war on socio-psychological well-being and social capital are long-lasting: even after the war ends, individuals who experienced greater violence exposure exhibit more retributive attitudes and deterioration in social capital, such as trust, norms, networks, and interpersonal relations. War impacts socio-psychological well-being and social capital by deteriorating the mental health of individuals. War exposure is associated with extremely high rates of depression, anxiety, stress-related mental illness, and severe psychiatric disorders (Vesco et al., 2025).

It may be possible to replenish the physical losses and rebuild the economy, but the more challenging aspect is that it takes a long time to repair the impact of war on people, and in some cases, it is not possible (Vesco et al., 2025). The echoes of war can be heard in society for generations to come and have multigenerational impacts of massive trauma (Yael Danieli, 2015).

III. Parental Experiences and Practices

In addition to standard war traumas, women are subject to gender specific violence. Women are at particularly high risk of experiencing acts of violence such as sexual violence, including rape, assault, trafficking, and prostitution, during many conflicts due to the breakdown of traditional safety structures (Devakumar et al., 2014). Females, at a rate of 49%, experienced at least one act of physical or sexual violence by a soldier or fighter during the First Liberian Civil War from 1989 through 1994 (Swiss, 1998). Exposure to war negatively reflects on a mother's mental health and contributes to the development of PTSD, depression, and anxiety, harsh punishment, parental rejection, and negative parental behavior. Moreover, a mother's psychopathology is directly related to their child's psychosocial problems, even after accounting for children's war trauma exposure (Sim et al., 2018). This can significantly affect the mental health of both mothers and children. Recent studies show that there are consistent effects of maternal mental health on developmental outcomes in child mental health.

To improve the mental health and psychosocial resilience of war-affected children, interventions should focus on psychosocial and parenting support for war-affected caregivers, as well as address structural challenges that debilitate caregivers and child mental health.

IV. Children's Physical and Mental Health Outcomes

Physical outcomes

The health of children is affected in many ways during armed conflict through injuries, acts of violence, diseases, malnutrition, forced migration, child soldiers, lack of education, loss of parents and socio-cultural changes, and many others. Millions of children are forced to carry the negative effects of war throughout their whole lives (Çelik & Özpınar, 2017).

i. Injuries.

Injuries to limbs and body extremities, burns, blindness, and deafness are more common for children during war than gunshots. Moreover, explosions cause more deaths and life-threatening injuries in children than in adults. Studies show that children are particularly vulnerable to blast injuries and burns, likely due to their smaller stature, physical vulnerability, and increased proximity to ground-level heat and blast effects. In December 2024, UNICEF reported that 20% of children in Ukraine require ongoing healthcare support, with explosive weapons causing 87% of casualties among children. Analysis of the war-related children's injuries during the Russo–Ukrainian War indicates that the most affected body areas included the upper extremities (31%), lower extremities (28%), and thorax (28%). This study findings are similar to studies from Afghanistan, Syria, and Iraq, although conflicts in the Middle East have reported a higher level of pelvic, head, and burn injuries, which were also the leading cause of pediatric mortality (Sennersten et al., 2025; Creamer et al., 2009). Unfortunately, large numbers of injured children remain physically handicapped for life because they are unable to receive healthcare in a timely manner, and rehabilitation services, like physical therapy, remain unavailable due to a lack of healthcare providers in war-affected areas. Based on studies completed in Afghanistan and Iraq, it is estimated that 40% of the children have suffered permanent disability from war-related injuries (Creamer et al., 2009).

ii. Diseases.

Exposure to armed conflicts is associated with the increased vulnerability of children against infectious, communicable, and noncommunicable diseases. Disruption of healthcare systems and lack of qualified personnel and medical services also contribute to the spread of infections and diseases in conflict-affected areas. The unavailability of immunization during armed conflicts, mass movements during the displacement and hardships of collective life in the refugee camps, make children susceptible to death from communicable and noncommunicable diseases, especially combined with malnutrition (Çelik & Özpınar, 2017). Children affected by armed conflict are less likely to receive vaccinations, which can lead to the spread of vaccine-preventable diseases during the time of combat and after the hostilities end (Novo et al., 2009).

The Yugoslav Wars are good examples of a time when a disease outbreak occurred due to gaps in immunization programs, contributing to the ongoing outbreak of polio, measles, mumps, and rubella (Novo et al., 2009). A study of the effects of the Persian Gulf War on infants and children reported that easily preventable and curable conditions such as diarrhea, acute respiratory infections, and fever are the common causes of death among children in the area of conflict, especially for children under five years of age (Ascherio et al., 1992). This could have been prevented if children had had access to healthcare professionals.

iii. Malnutrition.

In addition to the overall global food insecurity crisis (World Food Programme (WFP), n.d.), a huge portion of the war-affected population is suffering from malnutrition. The term ‘malnutrition is used to describe a deficiency, excess, or imbalance of a wide range of nutrients, resulting in measurable negative effects on body composition, function, and clinical outcome. According to UNICEF, undernutrition makes children much more vulnerable to diseases due to micronutrient deficiencies, with around 45% of deaths worldwide among children under 5 years old linked to malnutrition (“When it Matters Most”). It has been reported that undernutrition in Gaza causes wasting, stunting, being underweight, and deficiencies in vitamins and minerals amongst the population, particularly children, due to their specific nutritional needs. Low weight-for-height and low height-for-age indicators are being reported amongst children of Gaza as a result of chronic or repetitive malnutrition (Ashour et al., 2025). According to the WHO report, 30% of Gaza children under the age of five are experiencing severe malnutrition, with iron deficiency anemia affecting over 50% of children of the same age. Due to inadequacies in the intake of vitamins and minerals, referred to as micronutrients, many young children in Gaza might not be able to produce enzymes, hormones, and other substances that are essential for proper growth and development. Low muscle and bone density was also reported (Buheji et al., 2024). As of August 2025, 41.8% of pregnant women and 47.1% of breastfeeding women in Gaza were diagnosed with acute malnutrition, indicating the inability of mothers to provide proper nourishment for their children, leading to low birth weight and infant growth (“Nearly 1 in 2 Breastfeeding Mothers in Gaza Are Malnourished”).

All evidence suggests that chronic or repetitive malnutrition, especially in the early days of childhood, holds children back from reaching their physical and cognitive potential, and makes them vulnerable to infectious, communicable, and noncommunicable diseases. Without intervention, this can lead to long-term effects on health, development, or life opportunities.

Psychological outcomes.

Although basic survival necessities take priority over psychological needs in times of war and immediate crisis, there is growing concern that such overwhelming experiences will have a critically negative impact on the development of war-affected children, their mental health, attitudes towards society, relationships, and outlook on life in general. PTSD and depression are

the most commonly reported mental health disorders within war-affected children, as well as adults (Steel et al., 2009). The prevalence amongst children with direct war exposure for PTSD is assessed at 47%, depression at 43%, and anxiety at 27%. It is important to note that rates are lower amongst children with indirect exposures (Attanayake et al., 2009). Additional effects include behavioral and emotional symptoms, adversities, sleep problems, disturbed play, social hostility, memory and concentration problems, as well as many others (Erabi et al., 1975; Elvevåg & DeLisi, 2022). In several instances, children simultaneously experience many psychological adversities. For example, refugee children with PTSD frequently suffer from emotional distress and hyperactivity (Rizkalla et al., 2020).

Children who are living in conflict zones often exhibit externalizing behaviors such as aggression, irritability, defiance, and impulsiveness. These behaviors and strategies are coping mechanisms in response to overwhelming emotional stress. While these behaviors are often misunderstood and interpreted as misconduct, they are manifestations of trauma-induced difficulties in managing emotional responses and behaviors. In addition to externalizing symptoms, internalizing problems such as anxiety and depression are commonly reported, particularly among children exposed to chronic stress (Slone & Mann, 2016).

Unfortunately, the effects of war on children are not restricted to isolated mental health diagnoses but include a broad and multilayered set of developmental outcomes that compromise not just the child's health, but also family dynamics, social relations, academic achievement, and general life functions and satisfaction (Catani, 2018). Research suggests that prolonged stress and instability in children limit opportunities for healthy peer-to-peer interactions, structured play, and social learning, which leads to difficulty with the development of social skills, empathy, emotional regulation, and a general hopeful outlook on life (Rizkalla et al., 2020).

At the neurobiological level, chronic exposure to stress negatively effects regulatory systems such as the hypothalamic–pituitary–adrenal (HPA) axis. The HPA axis is a hormone-based communication system between the human brain and adrenal glands that regulates stress, energy, and mood. Deficiency of this system caused by prolonged stress or trauma can lead to fatigue, PTSD, anxiety, depression, memory loss, attention deficit, language acquisition, and immune system issues (Saad et al., 2025). Moreover, alcohol dependence is closely linked to abnormal cortisol levels. Coincidentally, it was determined by multiple studies that adolescents with cumulative exposure to war events and those with PTSD resulting from war events typically have significantly higher rates of alcohol and substance abuse (Zasiekina et al., 2025).

The psychological impact of war on children not only affects individual mental health but also places a substantial burden on families, schools, and healthcare systems. Family dynamics and parenting practices play a crucial role in children's psychological health during times of war; however, both could carry a risk and a protective factor depending on the situation. Studies have

shown that more war-affected parents show less warmth and more harshness toward their children (Eltanamy et al., 2021d). Parents with mental health problems may become irritable, viewing their child's mental health demands as highly stressful, which further contributes to the child's mental health problems.

While talking about the negative psychological effects on children, it is important to note that traumatic war experiences during childhood cannot determine the outcome of the developmental process (Murphy et al., 2022; Enlow et al., 2012). Even though these negative experiences carry a large impact on development, the effects of those experiences and trauma can be reversed. For example, a study of a family story time intervention among newly resettled Syrian refugees in Canada showed a significant decrease in anxiety symptoms in young children (Abi Zeid Daou et al., 2022). This promising revelation shows that timely positive intervention may be able to strengthen child development and reduce the effects of war.

Child soldiers.

Unfortunately, children are not just witnesses or passive victims of armed conflicts; in some situations, they become active participants, or child soldiers, with various armed forces and groups. Some children are forcibly abducted from their families and beaten into submission, some join military groups to escape extreme poverty and be among their peers, some join to defend their communities, or out of a feeling of revenge. According to Paris Principles and Commitments on Children Associated with Armed Forces or Armed Groups, a child soldier is a “person below 18 years of age associated with an armed force or armed group or who has been recruited or used by an armed force or group in any capacity, including but not limited to young boys and girls, used as fighters, cooks, porters, messengers, spies or for sexual purposes. It does not only refer to a child who is taking or has taken a direct part in hostilities.” It is important to note that even if a child perpetrated violence during the armed conflict, “children who are accused of crimes under international law allegedly committed while they were associated with armed forces or armed groups should be considered primarily as victims of offenses against international law; not only as perpetrators” (“The Paris Principles”). Child soldiers, first and foremost, are victims, whose participation in war has serious long-term physical and psychological consequences. In addition to common war-inflicted traumas and hardships, those children may have witnessed death, killing, torture, or sexual violence. They may have been subjected to serious abuse or forced to commit those actions themselves.

Many former child soldiers experience physical injuries and disabilities, high levels of anxiety, depression, PTSD, alcohol dependency, problems with social reintegration, stigma, and low community and family acceptance. It is also reported that former child soldiers are three times more likely to attempt suicide compared to non-child soldier war victims (Betancourt et al., 2020). Therefore, specialized and timely reintegration and rehabilitation care is extremely important for those children to ensure that they can rebuild their lives after they escaped armed

conflict. Medical care, psychosocial support, and education are important aspects of reintegration for former child soldiers, as well as reunion with their families and communities. It is essential to note that the successful rehabilitation of war-affected children is beneficial to civilization.

Evidence-based interventions for war-affected families

Parental emotional responses to war trauma and mental health are strongly linked to children's mental health and mental health disorders such as PTSD, anxiety, depression, and many others (Eltanamy et al., 2021). For example, it is stated by the US Department of Health and Human Services Office of the US Surgeon General that parental support has been associated with lower stress levels, better physical and mental health, better school performance, life satisfaction, and positive perceptions of life among children overall ("Parents Under Pressure"). It is suggested that supportive parent-child relationships are the driving factor that affects child outcomes in war-related traumas and might support healthy child adjustment (Enav & Mayer, 2025). Family members, particularly parents and children, influence each other through their adverse experiences; they also influence each other through their strengths and resilience in a time of crisis (Walsh, 2007). A powerful resource for war trauma treatment of children is the availability of family support during and following the traumatic event, and family bonding that builds a form of resilience through shared values and interdependence (Mak & Wieling, 2022). Therefore, healthcare providers recognize that caring for war-affected children means caring for the health needs of caregivers and families, facilitating access to care when necessary, and ensuring that children and their families have access to relevant, sustainable, and culturally-appropriate medical and social assistance.

Psychosocial interventions have been used successfully to complement psychiatric interventions in children and their families with diagnosed mental illnesses (Betancourt et al., 2013). Group sessions with parents on parenting skills, stress management techniques, problem-solving skills, self-care, and relapse prevention overall contribute to children's mental health and resilience indirectly, by improving parental well-being (Arega, 2023). Reinforcing positive everyday parental behavior, such as a parent hugging a child, can have a significant psychosocial impact on the recovery of the child (Mak & Wieling, 2022).

Studies conducted with children and their families who have experienced armed conflict stated that they look toward education as a tool to improve their lives and future. According to Save the Children UK, 38% of child refugees identified education as their first priority, and for 69% of children, education was among their three most highly prioritized sectors ("What do children want in times of emergency and crisis?"). Coordinated psychosocial and mental health interventions in school settings have been found to be beneficial and more socially and culturally acceptable for war-affected families. Education is especially relevant in cases where there is a cultural stigma associated with mental illnesses and a lack of awareness and understanding of mental health within refugee communities (Denov & C. Shevell, 2019)

There is evidence that the cultural competence of healthcare professionals has a positive outcome on patients' care. War-affected children and their families require care from professionals who are familiar with their unique health needs, and who are skilled in providing care to populations from different cultural and language backgrounds. For example, many refugee children and their families may not speak the language of their host country, making it difficult to communicate their needs and understand available resources. This language gap can lead to misunderstandings and misdiagnoses, further complicating the provision of appropriate care (Denov & C. Shevell, 2019).

According to UNICEF, psychosocial interventions that are aimed at establishing routine, creating safe and nurturing environments, addressing harmful behaviors, and engaging children and their families in structured activities, help families to gain a sense of control over their lives and cope with emotional trauma caused by war and conflicts ("Child protection"). In post-conflict environments, this may be accomplished by rebuilding homes, schools, and social and health care centers. In settings that are separated from direct conflict zones, such as asylum centers, this is accomplished by reuniting families, providing them with their own shelters, and enrolling children in school (Wessells, 2017).

Needs of war-affected children and families are complex and require integrated care models that address their unique psychological and practical needs (Arega, 2023). Support among international organizations, local governments, non-profit organizations, and local communities is crucial to ensure that war-affected children and their families receive the comprehensive support they deserve.

V. International response.

Armed conflicts do not only negatively affect children; they ultimately violate their rights that are clearly identified in the United Nations Convention on the Rights of the Child (UNCRC). The Convention on the Rights of the Child was adopted in 1989 by the United Nations, civil society, which involved countries reaching an agreement on a set of core rights for children (Heymann et al., 2014). The United Nations has identified six categories of human rights violations against children, known as the six grave violations. These violations include the killing and maiming of children, the abduction of children, the recruitment or use of children as soldiers, sexual violence against children, attacks against schools or hospitals, and the denial of humanitarian access. The first four are direct acts of violence against children, and the last two are indirect actions that cause harm to children and directly relate to health care and health workers. The commission of any of these violations constitutes a breach of international humanitarian law and carries legal consequences (UNICEF, 2009).

The United Nations Convention on the Rights of the Child is a legally binding international agreement that requires all participants to protect the civil, political, economic, social, and cultural rights of every child, irrespective of their race, religion, or abilities. Convention on the Rights of the Child (CRC) is built on four key principles, such as non-discrimination, right to life, survival and development, respect for views of the child, and best interests of the child. It also ensures that every child has a right to education, healthcare, protection from violence and exploitation, and the right to participate in cultural and social activities. While the UN provided a framework for protecting children's rights, all government entities and organizations involved must ensure children grow up in a safe and healthy environment (UN Article 6). A crucial feature of child rights is that they are entitlements that cannot be given or taken away by a government, and children's rights do not end at the border of war zones.

With rising global military expenditure worldwide and escalating conflict intensity, rivalries and geopolitical tensions must be managed amongst world powers (Report of the Secretary General on the impact of the global increase in military expenditure, 2024). It is critical that international organizations, governments, non-profit organizations, and local communities make the protection of children against armed conflicts their utmost priority. More than 473 million children are suffering from the effects of conflict worldwide. To put the issue into perspective, almost 19% of children live in conflict zones, according to UNICEF in 2024. Understanding the effects of war, displacement, and trauma on child development is crucial for developing well-designed interventions and support programs that fight those negative effects. Because children have little voice in shaping their future, global international organizations are required to fulfill the role of advocates for protection, promotion, and encouragement of participation in safeguarding the rights of war-affected children (Clark et al., 2020).

International organizations are instrumental in creating a stable and equitable global framework for promoting children's rights and well-being; however, even these organizations face challenges ("The Role of International Organizations in Global Public Policy Formation and Implementation"). While some organizations, such as UNICEF and Save the Children, promote international work across all nations and sectors, some organizations follow sectoral rather than holistic strategies and focus on specific areas of child well-being, such as diseases, education, malnutrition, protection from violence, and others. Sectoral divisions of the problems to make them manageable; however, it also suggests segmentation of the resources, information, an increase in bureaucracy, and limitations in response time (Clark et al., 2020).

While international organizations are continuously trying to advocate for the safety and well-being of children worldwide, compliance and execution of the Convention on the Rights of the Child depend on the leadership and commitment of governments. And yet, many politicians do not prioritize the well-being of children, nor see it as the foundation for broader societal

improvements and policymaking (Clark et al., 2020). Many studies advocate for countries to take deliberate, child-centered actions to coordinate and share responsibility for children across all sectors of government, establishing child well-being as a political priority and mobilizing domestic resources for children, developing overarching policies backed by national and international programs dedicated to child well-being, and aligning institutional incentives and accountability. Many governments dedicate specific government departments to oversee the upholding of children's rights; however, those departments often do not have the appropriate funding or political leverage to make any significant progress in expanding a child-centered agenda (Clark et al., 2020). Multiple countries, such as Iceland, Belgium, Nepal, Rwanda, South Africa, and a few others, have shown the ultimate legal commitment to the CRC by giving all or some of its provisions constitutional status (Rizky, L. 2022).

Despite many challenges, the Convention on the Rights of Children is the world's most ratified and accepted human rights treaty today. It offers a vision of the world in which all children are protected and respected. It demonstrates the power of children to unite people across nations for the common good of children and the future. International organizations and governments must focus on expanding the Convention on the Rights of Children framework and implement it across all sectors and areas of society to promote world peace, tolerance, fairness, and responsibility for human rights (UNICEF, 2009).

Conclusion.

Wars and armed conflicts have catastrophic effects on all aspects of children's lives, their families, and society. Effects of conflicts often include all six grave child rights violations established by the United Nations Security Council and children's rights indicated in the Convention on the Rights of Children. These effects also include a wide range of both direct and indirect physical and psychological chronic consequences that follow children through the course of life and into adulthood. This study explored a broad array of traumatic experiences that parents, caregivers, and families are exposed to during war. Additionally, the adverse effects that the parents' trauma and mental health can have on a child's development were also examined. As a consequence of their traumatic experiences, both parents and children are at an elevated risk of mental health issues, including post-traumatic stress disorder (PTSD), anxiety, depression, and many other mental health and psychosomatic symptoms. Studies suggest that family dynamics are commonly affected by war and conflicts, which, together with mental illness, have a negative impact on relations within the family and the psychological health of children. Coordinated and culture-appropriate psychosocial and mental health interventions, in concert with the incorporation of social institutions such as educational facilities, were proven to be effective in counteracting the negative effects of war in children and their families, and helping them to find their sense of safety and build resilience. Urgent actions are needed to protect millions of children who have been victims of war and continue to suffer from armed conflicts in many corners of the world. International organizations and government leadership must present a

united, organized, and child-centered solution to advance children's rights and child protection policies, compliance, and accountability, to counteract and eliminate the increasing rates of violence and frequency of armed conflicts.

References:

- Abi Zeid Daou, K. R., Abi Zeid Daou, L. R., & Cousineau-Pérusse, M. (2022). Storytelling as a tool: A family-based intervention for newly resettled Syrian refugee children. *International Journal of Social Welfare*, 31(1), 56–65. <https://doi.org/10.1111/ijsw.12476>
- Abuawad, A., Griffiths, M., Edwards, G. H., Eftekhari, A., Al-Ebweini, M., Al-Najar, H., Butmeh, A., Abu Dayyeh, R., El-Shewy, M., & Aker, A. (2025a). The Ongoing Environmental Destruction and Degradation of Gaza: The Resulting Public Health Crisis. *American Journal of Public Health*, 115(7), 1053–1061. <https://doi.org/10.2105/AJPH.2025.308140>
- Abuawad, A., Griffiths, M., Edwards, G. H., Eftekhari, A., Al-Ebweini, M., Al-Najar, H., Butmeh, A., Abu Dayyeh, R., El-Shewy, M., & Aker, A. (2025b). The Ongoing Environmental Destruction and Degradation of Gaza: The Resulting Public Health Crisis. *American Journal of Public Health*, 115(7), 1053–1061. <https://doi.org/10.2105/AJPH.2025.308140>
- Ahmad, A., Sofi, M. A., Sundelin-Wahlsten, V., & Von Knorring, A.-L. (2000). Posttraumatic stress disorder in children after the military operation “Anfal” in Iraqi Kurdistan. *European Child & Adolescent Psychiatry*, 9(4), 235–243. <https://doi.org/10.1007/s007870070026>
- Arega, N. T. (2023). Mental Health and Psychosocial Support Interventions for Children Affected by Armed Conflict in low-and middle-income Countries: A Systematic Review. *Child & Youth Care Forum*, 52(6), 1431–1456. <https://doi.org/10.1007/s10566-023-09741-0>
- Arsahanova, Z., Bokov, Y., Volgograd State University, Larin, A., & Russian State Academy of Intellectual Property. (2020). The impact of wars on the economy of countries: Theoretical and practical aspects. *Economic Annals-XXI*, 182(3–4), 34–40. <https://doi.org/10.21003/ea.V182-04>
- Ascherio, A., Chase, R., Coté, T., Dehaes, G., Hoskins, E., Laaouej, J., Passey, M., Qaderi, S., Shuqaidef, S., Smith, M. C., & Zaidi, S. (1992). Effect of the Gulf War on Infant and Child Mortality in Iraq. *New England Journal of Medicine*, 327(13), 931–936. <https://doi.org/10.1056/NEJM199209243271306>
- Attanayake, V., McKay, R., Joffres, M., Singh, S., Burkle, F., & Mills, E. (2009). Prevalence of mental disorders among children exposed to war: A systematic review of 7,920 children. *Medicine, Conflict and Survival*, 25(1), 4–19. <https://doi.org/10.1080/13623690802568913>
- Bendavid, E., Boerma, T., Akseer, N., Langer, A., Malembaka, E. B., Okiro, E. A., Wise, P. H., Heft-Neal, S., Black, R. E., Bhutta, Z. A., Bhutta, Z., Black, R., Blanchet, K., Boerma, T., Gaffey, M., Langer, A., Spiegel, P., Waldman, R., & Wise, P. (2021a). The effects of armed conflict on the health of women and children. *The Lancet*, 397(10273), 522–532. [https://doi.org/10.1016/S0140-6736\(21\)00131-8](https://doi.org/10.1016/S0140-6736(21)00131-8)

- Bendavid, E., Boerma, T., Akseer, N., Langer, A., Malembaka, E. B., Okiro, E. A., Wise, P. H., Heft-Neal, S., Black, R. E., Bhutta, Z. A., Bhutta, Z., Black, R., Blanchet, K., Boerma, T., Gaffey, M., Langer, A., Spiegel, P., Waldman, R., & Wise, P. (2021b). The effects of armed conflict on the health of women and children. *The Lancet*, 397(10273), 522–532. [https://doi.org/10.1016/S0140-6736\(21\)00131-8](https://doi.org/10.1016/S0140-6736(21)00131-8)
- Betancourt, T. S., Keegan, K., Farrar, J., & Brennan, R. T. (2020a). The intergenerational impact of war on mental health and psychosocial well-being: Lessons from the longitudinal study of war-affected youth in Sierra Leone. *Conflict and Health*, 14(1), 62. <https://doi.org/10.1186/s13031-020-00308-7>
- Betancourt, T. S., Keegan, K., Farrar, J., & Brennan, R. T. (2020b). The intergenerational impact of war on mental health and psychosocial well-being: Lessons from the longitudinal study of war-affected youth in Sierra Leone. *Conflict and Health*, 14(1), 62. <https://doi.org/10.1186/s13031-020-00308-7>
- Betancourt, T. S., McBain, R. K., Newnham, E. A., & Brennan, R. T. (2015). The intergenerational impact of war: Longitudinal relationships between caregiver and child mental health in postconflict Sierra Leone. *Journal of Child Psychology and Psychiatry*, 56(10), 1101–1107. <https://doi.org/10.1111/jcpp.12389>
- Betancourt, T. S., Meyers-Ohki, S. E., Charrow, A. P., & Tol, W. A. (2013). Interventions for Children Affected by War: An Ecological Perspective on Psychosocial Support and Mental Health Care. *Harvard Review of Psychiatry*, 21(2), 70–91. <https://doi.org/10.1097/HRP.0b013e318283bf8f>
- Betancourt, T. S., Thomson, D., & VanderWeele, T. J. (2018). War-Related Traumas and Mental Health Across Generations. *JAMA Psychiatry*, 75(1), 5. <https://doi.org/10.1001/jamapsychiatry.2017.3530>
- Bharati, T. (2022). The long shadow of the Kargil War: The effect of early-life stress on education. *Economics & Human Biology*, 44, 101097. <https://doi.org/10.1016/j.ehb.2021.101097>
- Bick, J., & Nelson, C. A. (2017). Early experience and brain development. *WIREs Cognitive Science*, 8(1–2), e1387. <https://doi.org/10.1002/wcs.1387>
- Blackmore, R., Gray, K. M., Boyle, J. A., Fazel, M., Ranasinha, S., Fitzgerald, G., Misso, M., & Gibson-Helm, M. (2020). Systematic Review and Meta-analysis: The Prevalence of Mental Illness in Child and Adolescent Refugees and Asylum Seekers. *Journal of the American Academy of Child & Adolescent Psychiatry*, 59(6), 705–714. <https://doi.org/10.1016/j.jaac.2019.11.011>
- Brück, T., Di Maio, M., & Miaari, S. H. (2019). Learning The Hard Way: The Effect of Violent Conflict on Student Academic Achievement. *Journal of the European Economic Association*, 17(5), 1502–1537. <https://doi.org/10.1093/jeea/jvy051>
- Buheji, M., Srividhya, & Hasan, A. (2024). *BEYOND FAMINE AND CHAOS - CASE OF GAZA*. <https://doi.org/10.17605/OSF.IO/B6SPQ>

- Castro-Vale, I., Severo, M., Carvalho, D., & Mota-Cardoso, R. (2019). Intergenerational transmission of war-related trauma assessed 40 years after exposure. *Annals of General Psychiatry, 18*(1), 14. <https://doi.org/10.1186/s12991-019-0238-2>
- Catani, C. (2018). Mental health of children living in war zones: A risk and protection perspective. *World Psychiatry, 17*(1), 104–105. <https://doi.org/10.1002/wps.20496>
- Çelik, N., & Özpınar, S. (2017a). Children and health effects of war Being a war child... *Cumhuriyet Medical Journal, 39*(4), 639–643. <https://doi.org/10.7197/223.v39i32356.368921>
- Çelik, N., & Özpınar, S. (2017b). Children and health effects of war Being a war child... *Cumhuriyet Medical Journal, 39*(4), 639–643. <https://doi.org/10.7197/223.v39i32356.368921>
- Clark, H., Coll-Seck, A. M., Banerjee, A., Peterson, S., DalGLISH, S. L., Ameratunga, S., Balabanova, D., Bhan, M. K., Bhutta, Z. A., Borrazzo, J., Claeson, M., Doherty, T., El-Jardali, F., George, A. S., Gichaga, A., Gram, L., Hipgrave, D. B., Kwamie, A., Meng, Q., ... Costello, A. (2020a). A future for the world’s children? A WHO–UNICEF–Lancet Commission. *The Lancet, 395*(10224), 605–658. [https://doi.org/10.1016/S0140-6736\(19\)32540-1](https://doi.org/10.1016/S0140-6736(19)32540-1)
- Clark, H., Coll-Seck, A. M., Banerjee, A., Peterson, S., DalGLISH, S. L., Ameratunga, S., Balabanova, D., Bhan, M. K., Bhutta, Z. A., Borrazzo, J., Claeson, M., Doherty, T., El-Jardali, F., George, A. S., Gichaga, A., Gram, L., Hipgrave, D. B., Kwamie, A., Meng, Q., ... Costello, A. (2020b). A future for the world’s children? A WHO–UNICEF–Lancet Commission. *The Lancet, 395*(10224), 605–658. [https://doi.org/10.1016/S0140-6736\(19\)32540-1](https://doi.org/10.1016/S0140-6736(19)32540-1)
- Committee on Improving the Health and Well-being of Children and Youth through Health Care System Transformation, Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education, Board on Health Care Services, Health and Medicine Division, & National Academies of Sciences, Engineering, and Medicine. (2024). *Launching Lifelong Health by Improving Health Care for Children, Youth, and Families* (T. L. Cheng & J. M. Perrin, Eds.; p. 27835). National Academies Press. <https://doi.org/10.17226/27835>
- Creamer, K. M., Edwards, M. J., Shields, C. H., Thompson, M. W., Yu, C. E., & Adelman, W. (2009). Pediatric Wartime Admissions to US Military Combat Support Hospitals in Afghanistan and Iraq: Learning from the First 2,000 Admissions. *Journal of Trauma: Injury, Infection & Critical Care, 67*(4), 762–768. <https://doi.org/10.1097/TA.0b013e31818b1e15>
- Denov, M., & C. Shevell, M. (2019). Social work practice with war-affected children and families: The importance of family, culture, arts, and participatory approaches. *Journal of Family Social Work, 22*(1), 1–16. <https://doi.org/10.1080/10522158.2019.1546809>

- Devakumar, D., Birch, M., Osrin, D., Sondorp, E., & Wells, J. C. (2014). The intergenerational effects of war on the health of children. *BMC Medicine*, 12(1), 57.
<https://doi.org/10.1186/1741-7015-12-57>
- Eltanamly, H., Leijten, P., Jak, S., & Overbeek, G. (2021a). Parenting in Times of War: A Meta-Analysis and Qualitative Synthesis of War Exposure, Parenting, and Child Adjustment. *Trauma, Violence, & Abuse*, 22(1), 147–160. <https://doi.org/10.1177/1524838019833001>
- Eltanamly, H., Leijten, P., Jak, S., & Overbeek, G. (2021b). Parenting in Times of War: A Meta-Analysis and Qualitative Synthesis of War Exposure, Parenting, and Child Adjustment. *Trauma, Violence, & Abuse*, 22(1), 147–160. <https://doi.org/10.1177/1524838019833001>
- Eltanamly, H., Leijten, P., Jak, S., & Overbeek, G. (2021c). Parenting in Times of War: A Meta-Analysis and Qualitative Synthesis of War Exposure, Parenting, and Child Adjustment. *Trauma, Violence, & Abuse*, 22(1), 147–160. <https://doi.org/10.1177/1524838019833001>
- Eltanamly, H., Leijten, P., Jak, S., & Overbeek, G. (2021d). Parenting in Times of War: A Meta-Analysis and Qualitative Synthesis of War Exposure, Parenting, and Child Adjustment. *Trauma, Violence, & Abuse*, 22(1), 147–160. <https://doi.org/10.1177/1524838019833001>
- Eltanamly, H., Leijten, P., Jak, S., & Overbeek, G. (2021e). Parenting in Times of War: A Meta-Analysis and Qualitative Synthesis of War Exposure, Parenting, and Child Adjustment. *Trauma, Violence, & Abuse*, 22(1), 147–160. <https://doi.org/10.1177/1524838019833001>
- Elvevåg, B., & DeLisi, L. E. (2022a). The mental health consequences on children of the war in Ukraine: A commentary. *Psychiatry Research*, 317, 114798.
<https://doi.org/10.1016/j.psychres.2022.114798>
- Elvevåg, B., & DeLisi, L. E. (2022b). The mental health consequences on children of the war in Ukraine: A commentary. *Psychiatry Research*, 317, 114798.
<https://doi.org/10.1016/j.psychres.2022.114798>
- Elvevåg, B., & DeLisi, L. E. (2022c). The mental health consequences on children of the war in Ukraine: A commentary. *Psychiatry Research*, 317, 114798.
<https://doi.org/10.1016/j.psychres.2022.114798>
- Enav, Y., & Mayer, Y. (2025a). Parenting in Armed Political Conflict: Relationships Between Parental Efficacy, Reflective Functioning, and Age Among War-Displaced Families. *Journal of Advanced Nursing*, jan.17034. <https://doi.org/10.1111/jan.17034>
- Enav, Y., & Mayer, Y. (2025b). Parenting in Armed Political Conflict: Relationships Between Parental Efficacy, Reflective Functioning, and Age Among War-Displaced Families. *Journal of Advanced Nursing*, jan.17034. <https://doi.org/10.1111/jan.17034>
- Erabi, T., Higuti, T., Kakuno, T., Yamashita, J., Tanaka, M., & Horio, T. (1975). Polarographic studies on ubiquinone-10 and rhodoquinone bound with chromatophores from *Rhodospirillum rubrum*. *Journal of Biochemistry*, 78(4), 795–801.
<https://doi.org/10.1093/oxfordjournals.jbchem.a130968>
- Fazel, M., Wheeler, J., & Danesh, J. (2005). Prevalence of serious mental disorder in 7000 refugees resettled in western countries: A systematic review. *The Lancet*, 365(9467), 1309–1314. [https://doi.org/10.1016/S0140-6736\(05\)61027-6](https://doi.org/10.1016/S0140-6736(05)61027-6)

- Ferrara, P., Cammisa, I., Zona, M., Corsello, G., Giardino, I., Vural, M., Pastore, M., Bali, D., & Pettoello-Mantovani, M. (2024). The Global Issue of Violence Toward Children in the Context of War. *The Journal of Pediatrics*, 269, 114007. <https://doi.org/10.1016/j.jpeds.2024.114007>
- Ferrara, P., Zona, M., Giardino, I., Scaltrito, F., Pettoello-Mantovani, C., & Cammisa, I. (2025a). The silent wounds of war: Psycho-physical impacts and international legal implications for children in conflict zones. *Global Pediatrics*, 14, 100287. <https://doi.org/10.1016/j.gpeds.2025.100287>
- Ferrara, P., Zona, M., Giardino, I., Scaltrito, F., Pettoello-Mantovani, C., & Cammisa, I. (2025b). The silent wounds of war: Psycho-physical impacts and international legal implications for children in conflict zones. *Global Pediatrics*, 14, 100287. <https://doi.org/10.1016/j.gpeds.2025.100287>
- Gredebäck, G., Hall, J., & Lindskog, M. (2022a). Fluid intelligence in refugee children. A cross-sectional study of potential risk and resilience factors among Syrian refugee children and their parents. *Intelligence*, 94, 101684. <https://doi.org/10.1016/j.intell.2022.101684>
- Gredebäck, G., Hall, J., & Lindskog, M. (2022b). Fluid intelligence in refugee children. A cross-sectional study of potential risk and resilience factors among Syrian refugee children and their parents. *Intelligence*, 94, 101684. <https://doi.org/10.1016/j.intell.2022.101684>
- Hazer, L., & Gredebäck, G. (2023a). The effects of war, displacement, and trauma on child development. *Humanities and Social Sciences Communications*, 10(1), 909. <https://doi.org/10.1057/s41599-023-02438-8>
- Hazer, L., & Gredebäck, G. (2023b). The effects of war, displacement, and trauma on child development. *Humanities and Social Sciences Communications*, 10(1), 909. <https://doi.org/10.1057/s41599-023-02438-8>
- Hazer, L., & Gredebäck, G. (2023c). The effects of war, displacement, and trauma on child development. *Humanities and Social Sciences Communications*, 10(1), 909. <https://doi.org/10.1057/s41599-023-02438-8>
- Heymann, J., McNeill, K., & Raub, A. (2014). Assessing Compliance with the crc. *The International Journal of Children's Rights*, 22(3), 425–445. <https://doi.org/10.1163/15718182-02203001>
- Kandemir, H., Karataş, H., Çeri, V., Solmaz, F., Kandemir, S. B., & Solmaz, A. (2018). Prevalence of war-related adverse events, depression and anxiety among Syrian refugee children settled in Turkey. *European Child & Adolescent Psychiatry*, 27(11), 1513–1517. <https://doi.org/10.1007/s00787-018-1178-0>
- Karatzias, T., Shevlin, M., Ben-Ezra, M., McElroy, E., Redican, E., Vang, M. L., Cloitre, M., Ho, G. W. K., Lorberg, B., Martsenkovskyi, D., & Hyland, P. (2023). War exposure, posttraumatic stress disorder, and complex posttraumatic stress disorder among parents living in Ukraine during the Russian war. *Acta Psychiatrica Scandinavica*, 147(3), 276–285. <https://doi.org/10.1111/acps.13529>

- Khamis, V. (2019a). Posttraumatic stress disorder and emotion dysregulation among Syrian refugee children and adolescents resettled in Lebanon and Jordan. *Child Abuse & Neglect*, 89, 29–39. <https://doi.org/10.1016/j.chiabu.2018.12.013>
- Khamis, V. (2019b). Posttraumatic stress disorder and emotion dysregulation among Syrian refugee children and adolescents resettled in Lebanon and Jordan. *Child Abuse & Neglect*, 89, 29–39. <https://doi.org/10.1016/j.chiabu.2018.12.013>
- Khamis, V. (2019c). Posttraumatic stress disorder and emotion dysregulation among Syrian refugee children and adolescents resettled in Lebanon and Jordan. *Child Abuse & Neglect*, 89, 29–39. <https://doi.org/10.1016/j.chiabu.2018.12.013>
- Kiser, L. J., Nurse, W., Lucksted, A., & Collins, K. S. (2008). Understanding the impact of trauma on family life from the viewpoint of female caregivers living in urban poverty. *Traumatology*, 14(3), 77–90. <https://doi.org/10.1177/1534765608320329>
- Kotsis, K. T. (2024). The Impact of War on the Environment. *European Journal of Ecology, Biology and Agriculture*, 1(5), 89–100. [https://doi.org/10.59324/ejeba.2024.1\(5\).07](https://doi.org/10.59324/ejeba.2024.1(5).07)
- Kuterovac, G., Dyregrov, A., & Stuvland, R. (1994). Children in war: A silent majority under stress. *British Journal of Medical Psychology*, 67(4), 363–375. <https://doi.org/10.1111/j.2044-8341.1994.tb01804.x>
- Leal Filho, W., Fedoruk, M., Paulino Pires Eustachio, J. H., Barbir, J., Lisovska, T., Lingos, A., & Baars, C. (2023). How the War in Ukraine Affects Food Security. *Foods*, 12(21), 3996. <https://doi.org/10.3390/foods12213996>
- Levy, B. S. (2025). The impacts of war on health, human rights, and the environment—An overview. *Frontiers in Public Health*, 13, 1547784. <https://doi.org/10.3389/fpubh.2025.1547784>
- Lynch, M. F. (2012). Theoretical Contexts of Trauma Counseling. In L. L. Levers (Ed.), *Trauma Counseling* (1st ed.). Springer Publishing Company. <https://doi.org/10.1891/9780826106841.0003>
- Mak, C., & Wieling, E. (2022). A Systematic Review of Evidence-Based Family Interventions for Trauma-Affected Refugees. *International Journal of Environmental Research and Public Health*, 19(15), 9361. <https://doi.org/10.3390/ijerph19159361>
- Martini, M., Massaro, E., Parrella, R., & Orsini, D. (2024). War and Health: The devastating impact of conflict on Well-being and Humanitarian Crises. *Journal of Preventive Medicine and Hygiene*, E464 Pages. <https://doi.org/10.15167/2421-4248/JPMH2024.65.3.3412>
- Mattelin, E., Paidar, K., Söderlind, N., Fröberg, F., & Korhonen, L. (2024). A systematic review of studies on resilience and risk and protective factors for health among refugee children in Nordic countries. *European Child & Adolescent Psychiatry*, 33(3), 667–700. <https://doi.org/10.1007/s00787-022-01975-y>
- Meaza, H., Ghebreyohannes, T., Nyssen, J., Tesfamariam, Z., Demissie, B., Poesen, J., Gebrehiwot, M., Weldemichel, T. G., Deckers, S., Gidey, D. G., & Vanmaercke, M. (2024). Managing the environmental impacts of war: What can be learned from conflict-

- vulnerable communities? *Science of The Total Environment*, 927, 171974.
<https://doi.org/10.1016/j.scitotenv.2024.171974>
- Murphy, F., Nasa, A., Cullinane, D., Raajakesary, K., Gazzaz, A., Sooknarine, V., Haines, M., Roman, E., Kelly, L., O'Neill, A., Cannon, M., & Roddy, D. W. (2022). Childhood Trauma, the HPA Axis and Psychiatric Illnesses: A Targeted Literature Synthesis. *Frontiers in Psychiatry*, 13, 748372. <https://doi.org/10.3389/fpsy.2022.748372>
- Negriff, S. (2020). Expanding our understanding of intergenerational exposure to adversity. *Children and Youth Services Review*, 118, 105369.
<https://doi.org/10.1016/j.childyouth.2020.105369>
- Novo, A., Huebschen, J. M., Muller, C. P., Tesanovic, M., & Bojanic, J. (2009a). Ongoing rubella outbreak in Bosnia and Herzegovina, March-July 2009—Preliminary report. *Eurosurveillance*, 14(39). <https://doi.org/10.2807/ese.14.39.19343-en>
- Novo, A., Huebschen, J. M., Muller, C. P., Tesanovic, M., & Bojanic, J. (2009b). Ongoing rubella outbreak in Bosnia and Herzegovina, March-July 2009—Preliminary report. *Eurosurveillance*, 14(39). <https://doi.org/10.2807/ese.14.39.19343-en>
- Panter-Brick, C., Dajani, R., Eggerman, M., Hermosilla, S., Sancilio, A., & Ager, A. (2018). Insecurity, distress and mental health: Experimental and randomized controlled trials of a psychosocial intervention for youth affected by the Syrian crisis. *Journal of Child Psychology and Psychiatry*, 59(5), 523–541. <https://doi.org/10.1111/jcpp.12832>
- Papageorgiou, V., Frangou-Garunovic, A., Iordanidou, R., Yule, W., Smith, P., & Vostanis, P. (2000). War trauma and psychopathology in Bosnian refugee children. *European Child & Adolescent Psychiatry*, 9(2), 84–90. <https://doi.org/10.1007/s007870050002>
- Peltonen, K. (2025a). Children and war – vulnerability and resilience. *European Journal of Developmental Psychology*, 22(3), 340–352.
<https://doi.org/10.1080/17405629.2024.2382410>
- Peltonen, K. (2025b). Children and war – vulnerability and resilience. *European Journal of Developmental Psychology*, 22(3), 340–352.
<https://doi.org/10.1080/17405629.2024.2382410>
- Qouta, S. R., Vänskä, M., Diab, S. Y., & Punamäki, R.-L. (2021). War trauma and infant motor, cognitive, and socioemotional development: Maternal mental health and dyadic interaction as explanatory processes. *Infant Behavior and Development*, 63, 101532.
<https://doi.org/10.1016/j.infbeh.2021.101532>
- Rizkalla, N., Mallat, N. K., Arafa, R., Adi, S., Soudi, L., & Segal, S. P. (2020a). “Children Are Not Children Anymore; They Are a Lost Generation”: Adverse Physical and Mental Health Consequences on Syrian Refugee Children. *International Journal of Environmental Research and Public Health*, 17(22), 8378.
<https://doi.org/10.3390/ijerph17228378>
- Rizkalla, N., Mallat, N. K., Arafa, R., Adi, S., Soudi, L., & Segal, S. P. (2020b). “Children Are Not Children Anymore; They Are a Lost Generation”: Adverse Physical and Mental Health Consequences on Syrian Refugee Children. *International Journal of*

- Environmental Research and Public Health*, 17(22), 8378.
<https://doi.org/10.3390/ijerph17228378>
- Rizkalla, N., Mallat, N. K., Arafa, R., Adi, S., Soudi, L., & Segal, S. P. (2020c). “Children Are Not Children Anymore; They Are a Lost Generation”: Adverse Physical and Mental Health Consequences on Syrian Refugee Children. *International Journal of Environmental Research and Public Health*, 17(22), 8378.
<https://doi.org/10.3390/ijerph17228378>
- Saad, H. A., Marzouk, M., Abdelrahman, H., & Moradikor, N. (2025). Mechanisms underlying stress effects on the brain: Basic concepts and clinical implications. In *Progress in Brain Research* (Vol. 291, pp. 21–47). Elsevier. <https://doi.org/10.1016/bs.pbr.2025.01.008>
- Samara, M., Hammuda, S., Vostanis, P., El-Khodary, B., & Al-Dewik, N. (2020a). Children’s prolonged exposure to the toxic stress of war trauma in the Middle East. *BMJ*, m3155.
<https://doi.org/10.1136/bmj.m3155>
- Samara, M., Hammuda, S., Vostanis, P., El-Khodary, B., & Al-Dewik, N. (2020b). Children’s prolonged exposure to the toxic stress of war trauma in the Middle East. *BMJ*, m3155.
<https://doi.org/10.1136/bmj.m3155>
- Sennersten, F., Frogh, S., Falk, G., Matvieienko, M., Karafulidi, O., Konstantynovska, O., Robinson, Y., Wladis, A., & Bäckström, D. (2025). War’s youngest victims: A descriptive cross-sectional study on injury distributions, severity patterns, and outcomes among paediatric trauma patients in Kharkiv, Ukraine. *Conflict and Health*, 19(1), 50.
<https://doi.org/10.1186/s13031-025-00694-w>
- Siehl, S., Sicorello, M., Herzog, J., Nees, F., Kleindienst, N., Bohus, M., Müller-Engelmann, M., Steil, R., Priebe, K., Schmahl, C., & Flor, H. (2022). Neurostructural associations with traumatic experiences during child- and adulthood. *Translational Psychiatry*, 12(1), 515.
<https://doi.org/10.1038/s41398-022-02262-9>
- Sim, A., Bowes, L., & Gardner, F. (2018). Modeling the effects of war exposure and daily stressors on maternal mental health, parenting, and child psychosocial adjustment: A cross-sectional study with Syrian refugees in Lebanon. *Global Mental Health*, 5, e40.
<https://doi.org/10.1017/gmh.2018.33>
- Slone, M., & Mann, S. (2016). Effects of War, Terrorism and Armed Conflict on Young Children: A Systematic Review. *Child Psychiatry & Human Development*, 47(6), 950–965.
<https://doi.org/10.1007/s10578-016-0626-7>
- Steel, Z., Chey, T., Silove, D., Marnane, C., Bryant, R. A., & Van Ommeren, M. (2009). Association of Torture and Other Potentially Traumatic Events With Mental Health Outcomes Among Populations Exposed to Mass Conflict and Displacement: A Systematic Review and Meta-analysis. *JAMA*, 302(5), 537.
<https://doi.org/10.1001/jama.2009.1132>
- Swiss, S. (1998). Violence Against Women During the Liberian Civil Conflict. *JAMA*, 279(8), 625. <https://doi.org/10.1001/jama.279.8.625>

- Szilagyi, P. G., & Schor, E. L. (1998). The health of children. *Health Services Research*, 33(4 Pt 2), 1001–1039.
- Thabet, A. A. M., Abed, Y., & Vostanis, P. (2002). Emotional problems in Palestinian children living in a war zone: A cross-sectional study. *The Lancet*, 359(9320), 1801–1804. [https://doi.org/10.1016/S0140-6736\(02\)08709-3](https://doi.org/10.1016/S0140-6736(02)08709-3)
- UNICEF (Ed.). (2009). *Celebrating 20 years of the Convention on the Rights of the Child*. UNICEF.
- Vesco, P., Baliki, G., Brück, T., Döring, S., Eriksson, A., Fjelde, H., Guha-Sapir, D., Hall, J., Knutsen, C. H., Leis, M. R., Mueller, H., Rauh, C., Rudolfsen, I., Swain, A., Timlick, A., Vassiliou, P. T. B., Von Schreeb, J., Von Uexkull, N., & Hegre, H. (2025a). The impacts of armed conflict on human development: A review of the literature. *World Development*, 187, 106806. <https://doi.org/10.1016/j.worlddev.2024.106806>
- Vesco, P., Baliki, G., Brück, T., Döring, S., Eriksson, A., Fjelde, H., Guha-Sapir, D., Hall, J., Knutsen, C. H., Leis, M. R., Mueller, H., Rauh, C., Rudolfsen, I., Swain, A., Timlick, A., Vassiliou, P. T. B., Von Schreeb, J., Von Uexkull, N., & Hegre, H. (2025b). The impacts of armed conflict on human development: A review of the literature. *World Development*, 187, 106806. <https://doi.org/10.1016/j.worlddev.2024.106806>
- Walsh, F. (2007). Traumatic Loss and Major Disasters: Strengthening Family and Community Resilience. *Family Process*, 46(2), 207–227. <https://doi.org/10.1111/j.1545-5300.2007.00205.x>
- Wessells, M. G. (2017). Children and armed conflict: Interventions for supporting war-affected children. *Peace and Conflict: Journal of Peace Psychology*, 23(1), 4–13. <https://doi.org/10.1037/pac0000227>
- World Food Programme (WFP). (n.d.). *WFP Annual Review 2024: Staying and Delivering Amid Multiple Crises*. World Food Programme (WFP). <https://doi.org/10.71958/WFP130793>
- Yayan, E. H. (2018). Post-traumatic stress disorder and mental health states of refugee children. *Archives of Psychiatric Nursing*, 32(6), 885–889. <https://doi.org/10.1016/j.apnu.2018.07.003>
- Zasiekina, L., Griffin, A., Blakemore, S.-J., Hlova, I., & Bignardi, G. (2025). Prevalence of war-related posttraumatic stress disorder in adolescents: A systematic review and meta-analysis. *European Journal of Psychotraumatology*, 16(1), 2497167. <https://doi.org/10.1080/20008066.2025.2497167>