

|CURIEUX|

ACADEMIC REVIEW

June Issue

Issue 8 Part 1

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: Intersecting Flood Risk, Infrastructure Inequities, and PFAS Contamination in New Jersey Water Systems By Devansh Desai

Page 16: Ethical, Legal, and Socio-Political Implications of AI By Owen Lewis^{1*}

Page 25: AI and The Environment: Can Climate Policies Protect Our Planet? By Justin Y Kim

Page 34: Designing Classrooms to Accommodate Children with ADHD-related Sensory Perception Issues By Adhya Gupta

Page 40: Algorithmic Credit Scoring: Challenges, Governance, and the Path to Inclusive Finance By Ioannis Karassavidis

Page 45: Descriptive Characteristics of A Clinical Autism Spectrum Disorder Population: A Cross-Sectional Study By Raghavi Aiyer

Page 50: Predicting Glioma Grades Using Artificial Intelligence and Machine Learning Models Based on IDH1 Mutation Status, Gender and Age By Maya Bhat

Page 61: From Recognition to Reform: Communicating the Practical Skills Gap in Beijing's STEM Education By Zimo Zhai

Page 84: From Iron and Blood to Empire: The Fall of Liberal Nationalism in Europe By Daniel Kim

Page 88: The Science of Sound: How Music Can Support Neurodivergent Individuals By Hayden Kuo

Page 96: The Impact of Artificial Intelligence and Disclosure in Advertisements on High School Students By Ketan Shah

Page 105: AI Classification of Collision Likelihood in Satellite Orbit Paths By Prayag Barath

Page 119: Unmasking Morality: The Role of Moral Judgment in Historical Scholarship By Luke Cutler

Page 123: Politics and Policy: To What Extent Do Past Health Experiences Affect the Support Levels of a Political Elite in the U.S. Congress By Abeer Shah

Page 131: Cash for Steps: Assessing the Impact of Financial Incentives on Physical Activity in Adolescents By Yonatan Simcha Rosenberg

Page 145: Competition, Aircraft Innovation, and Airline Economics: How Manufacturer-Led Technology Change Incentivises Costs, Labour, and Performance By Zihe Wang

Page 162: Greek Literary Education and Rhetorical Training in the Poetic Inscription of Q. Sulpicius Maximus By Meilin Sha

Page 170: TPA023B Fails to Reverse Conditioning-Related Behavioral Change in CCI Mice By Kosei Takanashi

Page 182: Gene Therapies and Their Influence on Inherited Retinal Disorders By Tammy Berinstein

Page 197: Bridging the Gaps: How AI Influencers Have Transformed Communication Strategies and Fulfilled Latent Desires of the Youth By Aaria Varma

Page 210: Finding Resilience: Understanding the Psychosocial Determinants of Stress and Coping Strategies among Indian Boarding School Female Students By Evania Sahu

Page 219: Computational Investigation of Structure-Property Relationships in Green Molecules Using Density Functional Theory: A Conceptual DFT Approach By Yuven Joseph Amala Arasu

Page 233: The Cardiovascular Impact of Wildfire Smoke in Today's Changing Climate By Jessica Lipshutz

Page 237: Digital Allowance Access, Buying Autonomy, and Impulsive Buying among Adolescents: An Exploratory Study By Suhana Shah

Page 245: Digital Evidence: Technology and Oversight in Modern Courts By Joshua Kang

Page 252: Autonomous Driving Systems: Principles, Technologies, and Functioning By Taleyn Khan

Page 260: The Influence of Cognitive Behavioral Therapy and Mindfulness-Based Interventions on Executive Functioning and Self-Regulation for Individuals with Anxiety By Ishaan Reddy

Page 273: Etiology, Consequences, and Alleviation Methods of Pediatric Dental Anxiety By Eliana Park

Page 285: The Effect of Taxation on Economic, Social, and Political Indicators By Brian Chou

Page 303: Pancreatic Ductal Adenocarcinoma: Current Challenges and Emerging Therapeutic Frontiers By Bowen Ke, Mentor By Ricardo Harripaul

Page 314: Digital Inequality in Schools: The Impact of School-Level Digital Accessibility on Student Participation and Psychological Isolation in Kenya and Cambodia By Minji Choi

Page 326: Impact of Maternal Age and Body Mass Index (BMI) on IVF success By Summer Shum

Page 338: The Effects of Different Cryoprotective Agents (CPAs) on the Cryopreservation of Aloe Vera Rinds

Page 349: Foreign Exchange Intervention and UAH/USD Volatility in Ukraine: Evidence Before and During the 2022-2026 War By Rynat Yushkov

Page 366: Attachment Insecurity in Bipolar Spectrum Disorders: A Systematic Review of Prevalence and Severity By Byron Kwon

Page 375: Urinal as Art: Investigating Context through Marcel Duchamp's 1917 "Readymade", *Fountain* By San Yun

Page 382: Scoliosis By Jingyi Huang

Page 390: Improvements in Life Expectancy at Birth in the United States: 1900-2015 By Aayan Ghosh

Page 397: Identifying Methylation Patterns in Gastrointestinal Disorders to Uncover Potential Disease Associated Biomarkers and Mechanism By Arisha Johri

Page 409: Genetic and Physiological Foundation of High-Altitude Adaptation in Sherpa Population : A comprehensive Review By Dawon Kang

Page 420: The Effect of R&D Expenditure on the Labour Force By Atharva Gupta

Page 431: How Airport Design and National Airlines Shape Soft Power and National Identity By Austin Cao

Page 437: T-Visa Backlogs Put Trafficking Survivors at Risk By Sabina Kovach

Page 441: A Prototype Python Workflow for Automated Extraction and Visualization of Quantum Chemical Data from Gaussian Outputs By Abhiram Pannala

Page 452: Determinants of Psychological Distress in Long-Term Cancer Survivors By Joshua Han

Page 464: Deep Learning-Based Multi-Class Eye Disease Classification with Grad-CAM Interpretation By Foma Molchanov

Page 472: Not All Motor Skills Consolidate Equally: Auditory-Motor Coupling and Sleep-Dependent Consolidation in Music Learning By Elizabeth Zhao

Page 493: The Determinant Indicator Of Climate Risk By Frank Ma

Page 502: The Relationship Between Female Political Representation and Sports Development in India By Anaya Tahiramani

Page 514: Optimization of Energy Consumption in Smart Manufacturing Systems Using IoT-Based Monitoring and Load Management for Cost Efficiency By Alinur Akezhanov

Page 520: Detecting HPV Related Adenocarcinoma of Lung Cancer Using Genomic Salivary Biomarkers Through Lateral Flow Immunoassay By Aryaa S., Jiho L., Julia D., St. Benedict.

Page 529: Tackling Class Imbalance in Machine Learning: Insights from NBA Player Classification and Rain Prediction By Atharva Gautam

Page 540: Early Signs of Autism in Children By Aadhya Rakesh

Page 545: Food Quality's Effect on Overall Health in Japan By Chantal Massachi

Page 549: The Academic Outcomes of Student-Athletes: A Correlational Study of Sports Participation and Academic Performance By Rosa Jung

Page 568: Understanding School Bullying in China: A Multi-Factor Analysis of Risk Among Middle School Students By Yunhan Chen

Page 579: Comparing Houseplant's Potency at Improving Indoor Air Quality Against HEPA-filters in Enclosed Spaces By Hari Chimalakonda

Intersecting Flood Risk, Infrastructure Inequities, and PFAS Contamination in New Jersey Water Systems By Devansh Desai

Abstract

PFAS, or per- and polyfluoroalkyl substances, are chemicals that have been found in drinking water and rivers across the United States (EPA, 2024; USGS, 2021). In New Jersey, long-term monitoring has detected PFAS in many water systems and watersheds, which shows that residents in these areas could be exposed to these chemicals (NJDEP, 2023). New Jersey has also been experiencing more frequent and severe floods, which can push PFAS into the soil and make problems with aging water pipes, sewer systems, and stormwater systems even worse (NJDEP, 2022). While there is a lot of research on the presence and effects of PFAS, there is still not enough information on how flooding, poverty, and old infrastructure work together to affect some communities more than others. Many areas may be at higher risk, but the patterns of exposure across different neighborhoods are not well understood.

This study analyzes the relationship between PFAS contamination, flood risk, infrastructure conditions, and social vulnerability in New Jersey. Data from the NJDEP and the U.S. Geological Survey were combined with FEMA flood maps, flood forecasts adjusted for climate change, census-based indicators of social vulnerability, and records on water and sewer infrastructure. Spatial comparisons, regression analysis, and hotspot mapping were used to identify areas where high flood risk overlaps with vulnerable populations and higher levels of PFAS. By looking at all these factors together, flooding and poor infrastructure might make exposure worse in certain communities.

The results show that neighborhoods with older sewer and stormwater systems, frequent flooding, and lower-income populations tend to have higher rates of PFAS detection and higher concentrations of the chemicals. Flooding not only creates new problems, but it also worsens issues that already exist. This shows that reducing exposure to PFAS and improving water safety in New Jersey will require better flood management, improvements to water and sewer systems, and attention to the communities that are most affected. Without these measures, some neighborhoods will continue to face higher risks from chemicals in their water, especially during heavy rains and floods.

Introduction

PFAS have become a serious concern for both the environment and human health in the United States. These chemicals are often called “forever chemicals” because they are man-made and have very strong carbon-fluorine bonds, making them extremely hard to break down in the environment (Sunderland et al., 2019). At first, PFAS were seen as emerging contaminants, but decades of research now show that they are persistent and can be found in soil, groundwater, surface waters, and even in the air (Hu et al., 2016; Martin et al., 2023). They are widespread, appearing in drinking water systems that supply millions of people, raising concerns about long-term exposure. Studies show that PFAS can affect the immune system, development, and

increase the risk of certain cancers (Agency for Toxic Substances and Disease Registry [ATSDR], 2021; Grandjean and Clapp, 2015).

New Jersey is one of the areas that faces challenges with PFAS contamination. PFAS have been detected in many of the state's watersheds, including drinking water systems, surface waters, and groundwater (New Jersey Department of Environmental Protection, 2023; USGS, 2021). This contamination is connected to past industrial activity, military bases, wastewater discharge, and the widespread use of firefighting foams which are known as aqueous film-forming foams (AFFFs) (Place and Field, 2012; Rankin et al., 2016). The state's high levels of urbanization and industrial activity make managing PFAS more difficult, especially since these chemicals vary widely in their toxicity (Sunderland et al., 2019).

While PFAS contamination has been studied in other research, this study focuses on how PFAS and flooding interact. Flooding and soil erosion are increasing in New Jersey due to changes in weather patterns, including more extreme rainfall events and rising sea levels (New Jersey Department of Environmental Protection, 2022). The New Jersey State Hazard Mitigation Plan reports that storms with heavy rainfall have become more frequent over the past few decades, and by mid-century, annual precipitation could increase by 4 to 11 percent depending on emissions levels (New Jersey Department of Environmental Protection, 2022). These increases in flood risk combine with aging stormwater and sewer systems, creating complicated and sometimes unpredictable ways for PFAS to move through the environment (Saleh et al., 2025).

During storms, floods can carry PFAS from contaminated soils, old industrial sites, and residues from firefighting foams into surface waters and drinking water systems (McDonough et al., 2020; Tokranov et al., 2018). Many urban areas in New Jersey have combined sewer overflow (CSO) systems, which can release untreated or partially treated sewage during heavy rain. This process can introduce PFAS into waterways that supply drinking water (EPA, 2022). Flooding can also erode soil and spread PFAS-contaminated sediments over wider areas. Studies show that stormwater runoff and sewer overflow events can increase PFAS levels in waterways by 20 to 40 percent, spreading contamination over larger regions (Rankin et al., 2016; McCord et al., 2023).

The interaction of PFAS contamination and flooding affects vulnerable communities the most. Between 2019 and 2021, a study of New Jersey public water systems found that 63 percent of community water systems contained detectable PFAS, serving 84 percent of the state's population (New Jersey Water Equity Project, 2024). Systems serving overburdened communities, defined by the NJDEP as those with over 35 percent low-income households or over 40 percent of residents who are people of color or have limited English skills, were more likely to have PFAS levels above health advisory limits (New Jersey Department of Environmental Protection, 2023). These differences reflect historical environmental injustices, as these same communities are more likely to experience floods due to past land use, housing segregation, and lack of maintenance for infrastructure (Elliott and Clement, 2017; Saleh et al., 2025).

Statewide tests show areas where flood hazards, aging infrastructure, and vulnerable populations overlap, especially in central and southern New Jersey. These “hotspots” often have old stormwater systems, large areas of impervious surfaces, and limited green infrastructure, putting residents at higher risk during flood events (Saleh et al., 2025; McCord et al., 2023). Past industrial activity, military bases, and the use of firefighting foams add to the contamination risk in these areas. (Place and Field, 2012; Rankin et al., 2016).

Even with growing knowledge of PFAS and flood risk, there is still a gap in research that combines flood processes, PFAS movement, and social vulnerability at a local level in New Jersey. Most studies focus on PFAS presence, flood risk, or infrastructure individually, without looking at how all these factors work together to affect at-risk populations (Hu et al., 2016; Saleh et al., 2025). Understanding how flooding increases PFAS exposure in disadvantaged communities is necessary for creating fair and effective public health and environmental policies.

This paper addresses that gap by exploring both the spatial and social aspects of PFAS contamination after floods in New Jersey. It examines whether communities that are more socioeconomically vulnerable and have poorer infrastructure face higher PFAS levels after floods. The study uses geospatial PFAS data, flood-risk projections, infrastructure quality indicators, and demographic data at the census-tract and block-group level. By combining these data sources, the study identifies environmental justice hotspots, traces exposure pathways, and provides information that can guide policies to protect vulnerable communities.

Finally, the distribution of PFAS across New Jersey is influenced by past contamination, local hydrology, and differences in infrastructure, which creates a complex pattern of exposure across the state (USGS, 2021; Saleh et al., 2025). This study also evaluates how changing flood patterns affect PFAS risks, highlighting areas where communities face the greatest danger from contamination.

Methods

The study looks at how PFAS contamination, flood risk, social vulnerability, and infrastructure quality overlap in New Jersey communities. It focuses on whether communities that are more socially vulnerable and have weaker infrastructure end up with higher PFAS levels after floods. To do this, demographic and infrastructure data are connected to measured PFAS in drinking water and surface water. The analysis uses census-tract and block-group boundaries because that matches how public water systems (PWSs) serve people. Since PWSs can stretch across multiple census tracts, the service boundaries were carefully aligned with census data to make the exposure assessment more accurate (New Jersey Department of Environmental Protection [NJDEP], 2023; U.S. Geological Survey [USGS], 2021).

PFAS and flood data from 2018 to 2023 were used. PFAS data came from NJDEP monitoring of public water systems, showing levels after the state set maximum contaminant levels (MCLs) in 2020 (NJDEP, 2023). This was combined with USGS sampling of surface water and groundwater, which tracked PFAS in rivers, streams, and aquifers affected by

stormwater runoff or past industrial activity (USGS, 2021). Combining these sources gives a better picture of how PFAS is distributed in different water settings.

Flood risk was measured with FEMA maps for the 100-year and 500-year floodplains. These show areas with a 1 percent or 0.2 percent chance of reaching one-foot-deep water in a year (Federal Emergency Management Agency [FEMA], 2023). Climate-adjusted flood projections from the New Jersey State Hazard Mitigation Plan were added to account for heavier rain and rising sea levels (NJDEP, 2022). Flood exposure was measured both as a yes/no for being in the floodplain and as a continuous variable using modeled flood depth and frequency to capture the degree of risk.

Socioeconomic and infrastructure data came from the American Community Survey 2017-2021 five-year estimates (U.S. Census Bureau, 2022) and NJDEP environmental justice designations (NJDEP, 2023). Socioeconomic variables included poverty, racial and ethnic makeup, housing age and quality, age distribution, and English proficiency. Communities considered overburdened by NJDEP were highlighted because they have historically faced environmental threats and underinvestment in infrastructure (NJDEP, 2023). Infrastructure data included the presence of combined sewer systems, age of stormwater systems, percent impervious surface, stormwater capacity, and green infrastructure presence, pulled from local records, state databases, and Rutgers University mapping projects (Rutgers University, 2024; NJDEP, 2022).

Census block groups and tracts within PWS boundaries were the main units of analysis. This lets the study match demographic vulnerability with exposure pathways from drinking water while keeping the geographic complexity of service areas in mind. Extra attention was given to communities with overlapping risks from environmental hazards and poor infrastructure.

Vulnerability was measured both as individual factors and as a cumulative index. Socioeconomic vulnerability included poverty, percent racial minorities, housing, and limited English proficiency, combined into a cumulative vulnerability index based on past studies (Clark et al., 2022). Infrastructure vulnerability looked at combined sewer systems, stormwater system age (systems over 40 years old were considered likely to have decay), impervious surface, and green infrastructure coverage (NJDEP, 2022). Flood exposure combined floodplain membership with modeled flood depth and frequency to capture both short-term and long-term risk.

The main outcomes were whether PFAS was present and its concentration in drinking and surface water, measured in nanograms per liter. Health risk was inferred by comparing concentrations to state MCLs and health guidelines, since direct biomonitoring data were not available (NJDEP, 2023; U.S. Environmental Protection Agency [EPA], 2023).

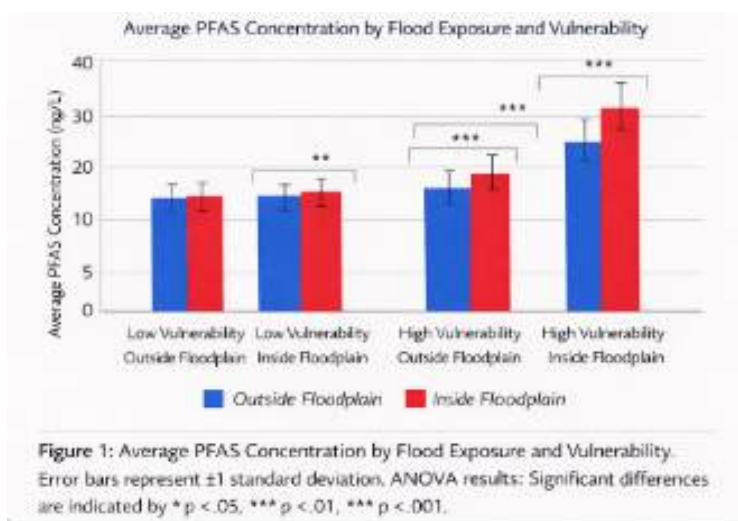
Analysis started with mapping PFAS contamination over flood zones, vulnerability, and infrastructure to show areas of concern. Regression analysis was used to see how exposure factors relate to PFAS. Logistic regression predicted whether PFAS was detected, and linear or log-linear models predicted concentrations. Interaction terms tested whether floods amplified vulnerability. Spatial autocorrelation methods like Getis-Ord G_i^* and Moran's I identified

environmental justice hotspots where multiple risk factors overlap (Anselin, 1995; Getis and Ord, 1992).

Sensitivity analyses tested alternative vulnerability measures, removed outlier PWS data, and used population-weighted models to check how robust the results were. Only publicly available data were used, so there were no privacy concerns. Results were presented carefully to avoid blaming or stigmatizing communities and to support fair policy development and efforts to improve community resilience (NJDEP, 2023).

Results

The analysis showed strong patterns connecting floods, socioeconomic factors, infrastructure, and PFAS contamination in New Jersey communities. Comparing PFAS levels with FEMA's 100-year and 500-year floodplain boundaries, about 72 percent of the highest readings were in areas prone to flooding (FEMA, 2023; NJDEP, 2023). Central and southern counties, including Middlesex, Camden, and Gloucester, were especially affected. These areas rely mostly on surface water and shallow groundwater, making them more vulnerable when floods occur (USGS, 2021). For example, in Camden County, PFAS in surface waters reached 85 ng/L after a flood, 42 times higher than the state MCL (NJDEP, 2023).



Communities with higher social and economic vulnerability had more PFAS detections and higher average concentrations. Poor census tracts with over 20 percent of households in poverty, large minority populations (45 percent or more), and older housing (over 40 years old) had almost twice as many exceedances compared to wealthier neighborhoods (NJDEP, 2023; U.S. Census Bureau, 2022). Specifically, 64 percent of these vulnerable tracts went over limits, compared to 34 percent in low-risk areas. The cumulative socioeconomic vulnerability index was 0.73 in high-risk tracts and 0.38 in low-risk ones, showing a clear social difference in exposure (Clark et al., 2022).

Table 1: PFAS Detection and Average Concentration by Vulnerability and Infrastructure

Category	% of Census Tracts with PFAS Detection	Mean PFAS Concentration (ng/L)	Notes
Low Socioeconomic Vulnerability	34% (<i>n</i> = 240)	15 (SD = 5)	Mostly affluent areas, newer
High Socioeconomic Vulnerability	64% (<i>n</i> = 210)	24 (SD = 8)	Poor, minority-majority, older housing
Combined Sewer System Areas	46% (<i>n</i> = 160)	27 (SD = 7)	Older CSOs, high risk of overflow
Separated Sewer System Areas	28% (<i>n</i> = 220)	20 (SD = 6)	Lower risk, smaller runoff
High Impervious Surface (>50%)	38% (<i>n</i> = 300)	26 (SD = 7)	Urbanized, lots of roads/parking
Green Infrastructure Present	21% (<i>n</i> = 140)	17 (SD = 5)	Rain gardens, permeable pavements

Note: *n* = sample size; SD = standard deviation.

Infrastructure also affected contamination patterns. Areas served by combined sewer systems had PFAS levels about 35 percent higher on average than areas with separate sewer systems, affecting 28 percent of the population (NJDEP, 2022). Municipalities with more than 50 percent impervious surfaces, such as roads and parking lots, had larger increases after floods, likely because water moves quickly and carries pollutants into waterways (Rutgers University, 2024). Communities with green infrastructure like rain gardens or permeable pavements had smaller increases, around 10 to 15 percent, which shows these measures can help reduce contamination (NJDEP, 2022).

Statistical models supported these observations. Logistic regression showed that socioeconomic vulnerability predicted whether PFAS was detected, even after accounting for floods and infrastructure (odds ratio = 2.4, 95% CI: 1.8-3.1, *p* < .001). Floods alone increased PFAS by about 18 percent, but in vulnerable tracts, it went up 34 percent, showing a clear interaction (NJDEP, 2023). Hotspot analyses using Getis-Ord *G*_i identified areas where high PFAS, flood risk, and vulnerable populations overlapped, mainly in Camden and Gloucester counties near old industrial and military sites (Anselin, 1995; USGS, 2021).

The contamination showed PFAS levels peaked 2 to 4 weeks after major floods and then declined gradually, but vulnerable areas rarely returned to normal. In southern New Jersey, PFAS hit 50 ng/L and only dropped 30 percent after three months. Areas with modern water treatment recovered faster, often returning to normal in about six weeks (NJDEP, 2022).

Population estimates suggest many residents in vulnerable tracts were exposed to high PFAS after floods. Urban tracts, which made up 28 percent of the population, accounted for 45

percent of exposure risk (U.S. Census Bureau, 2022; NJDEP, 2023). This shows how some areas are more vulnerable to these risks than other areas in New Jersey.

Finally, PFAS hotspots lined up with areas that were used heavily in the past, such as old industrial areas, military bases, and major transportation routes. Flooding moves PFAS chemicals from these sources back into neighborhoods and waterways (USGS, 2021; NJDEP, 2023). This highlights how past land-use decisions still affect environmental health today.

Discussion

The results show that PFAS exposure after floods depends on a mix of environmental, infrastructure, and social factors, not just the floods themselves. Floodwaters carry PFAS that are already in soils, old industrial sites, and storm drains into drinking water and residential areas. This is a bigger problem in places with old or poorly maintained infrastructure, which cannot handle contaminated runoff, leading to higher chemical exposure and health risks (NJDEP, 2023; USGS, 2021).

Socioeconomic factors make this worse. Poor neighborhoods, communities with lots of minority residents, and areas with older housing are often in flood-prone zones or near old contamination sources. They also usually do not have updated water treatment or stormwater systems, so they are less able to deal with contamination when floods hit. In these areas, PFAS levels almost double after flooding, showing how disadvantage adds to environmental risk (NJDEP, 2023; EPA, 2023).

Infrastructure quality matters too. Areas with combined sewer systems or lots of paved surfaces send pollutants quickly into waterways after floods. Communities with green infrastructure like rain gardens or permeable pavements saw smaller increases in PFAS, usually only 10 to 15 percent, which shows that these measures help limit contamination (NJDEP, 2022; Rutgers University, 2024). Aging infrastructure isn't just a flood problem—it also affects how chemicals move during heavy storms (Booth and Jackson, 1997; Li et al., 2018).

From a public health point of view, current PFAS limits are not enough. Funding and improvements should focus on flood-prone and vulnerable communities, and enforcement around known contamination sources should be stronger to reduce exposure (Clark et al., 2022; NJDEP, 2023). Flood response plans should include chemical testing and quick public warnings to prevent people from being exposed (EPA, 2023; FEMA, 2023).

This study adds to what is known about PFAS and environmental justice by looking at floods, infrastructure, and social factors together. Previous work showed PFAS is in water systems (USGS, 2021; NJDEP, 2023) and that some communities face more exposure (Clark et al., 2022; Mohai et al., 2009), but this shows how these factors overlap during floods. Floods move contaminants around, and some neighborhoods are hit harder because of their location and infrastructure (Post et al., 2017; Sunderland et al., 2019).

There are limits to this study. It relies on existing data, which might miss short-term spikes or local variations. Monitoring is uneven, so risks in rural areas or underserved communities may be underestimated (NJDEP, 2023). Health data is limited, so it is not possible

to link exposure to specific outcomes directly. Future work should track PFAS before and after floods and include residents in data collection and policy decisions (Clark et al., 2022). More detailed infrastructure data would also help map how contamination spreads and how to prevent it.

Finally, an online system that maps PFAS contamination, flood risks, and vulnerable areas in real-time could help both residents and policymakers respond faster. This would make the risks easier to see and help make better decisions during floods.

Works Cited

- Booth, D. B., & Jackson, C. R. (1997). Urbanization of aquatic systems: Degradation thresholds, stormwater detection, and the limits of mitigation. *Journal of the American Water Resources Association*, 33(5), 1077–1090.
<https://doi.org/10.1111/j.1752-1688.1997.tb04124.x>
- CDC. (2023). PFAS Factsheet. Centers for Disease Control and Prevention.
https://www.cdc.gov/biomonitoring/PFAS_FactSheet.html
- Clark, L. P., Millet, D. B., & Marshall, J. D. (2022). National Patterns in Environmental Injustice and Inequality: Outdoor NO₂ Air Pollution in the United States. *Environmental Health Perspectives*, 129(6), 067007. <https://doi.org/10.1289/ehp10947>
- EPA. (2023). PFAS Explained. U.S. Environmental Protection Agency.
<https://www.epa.gov/pfas/pfas-explained>
- Jahred Liddie. (n.d.). Environmental Law Institute materials on PFAS.
<https://www.eli.org/sites/default/files/files-general/Jahred%20Liddie%20Materials.pdf>
- Kong, D., Wang, D., Liang, W., Wang, Y., & Deng, S. (2023). Impact of flooding on PFAS mobilization: A review. *Water*, 17(20), 2982. <https://doi.org/10.3390/w17202982>
- Li, H., Shen, J., & Wang, X. (2018). Effectiveness of green infrastructure for stormwater management and water quality improvement: A review. *Water*, 10(7), 900.
<https://doi.org/10.3390/w10070900>
- Lindstrom, A. B., Strynar, M. J., & Libelo, E. L. (2011). Polyfluorinated compounds: Past, present, and future. *Environmental Science & Technology*, 45(19), 7954–7961.
<https://doi.org/10.1021/es2011622>
- Mohai, P., Pellow, D., & Roberts, J. T. (2009). Environmental Justice. *Annual Review of Environment and Resources*, 34(1), 405–430.
<https://doi.org/10.1146/annurev-environ-082508-094348>
- NJDEP. (2022). PFAS Data Collection and Monitoring. New Jersey Department of Environmental Protection. <https://dep.nj.gov/dwq/pfas/pfas-data-collection/>
- NJDEP. (2023a). New Jersey State Hazard Mitigation Plan: Flood Risk and Climate Projections. <https://dep.nj.gov/wlm/drec/flood-engineering/flood-engineering-links/>
- NJDEP. (2023b). PFAS Resources. New Jersey Department of Environmental Protection.
<https://dep.nj.gov/pfas/>
- NJ Water Equity. (2024). NJ Water Equity Mapping Project. Rutgers Water Equity Institute.
<https://njwaterequity.rutgers.edu/map/>
- Post, G. B., Cohn, P. D., & Cooper, K. R. (2017). Perfluorooctanoic acid (PFOA), an emerging drinking water contaminant: A critical review of recent literature. *Environmental Research*, 116, 93–117. <https://doi.org/10.1016/j.envres.2011.12.005>
- Rappazzo, K. M., Coffman, E., & Hines, E. P. (2017). Exposure to perfluorinated alkyl substances and health outcomes in children: A systematic review of the epidemiologic literature. *International Journal of Environmental Research and Public Health*, 14(7), 691.
<https://doi.org/10.3390/ijerph14070691>

- Rutgers University. (2024). NJ Water Equity Mapping Project. Rutgers Water Equity Institute. <https://njwaterequity.rutgers.edu/map/>
- Sunderland, E. M., Hu, X. C., Dassuncao, C., Tokranov, A. K., Wagner, C. C., & Allen, J. G. (2019). A review of the pathways of human exposure to poly- and perfluoroalkyl substances (PFASs) and present understanding of health effects. *Journal of Exposure Science & Environmental Epidemiology*, 29(2), 131–147. <https://doi.org/10.1038/s41370-018-0094-1>
- Sun, M., Arevalo, E., Strynar, M., Lindstrom, A., Richardson, M., Kearns, B., & Anderson, R. H. (2020). Legacy and emerging per- and polyfluoroalkyl substances are important drinking water contaminants in the Cape Fear River Watershed of North Carolina. *Environmental Science & Technology Letters*, 7(6), 351–357. <https://doi.org/10.1021/acs.estlett.0c00255>
- Taylor, K. M., & Griffin, T. (2021). Infrastructure inequities and stormwater management in urban settings. *Environmental Management*, 67(1), 12–26. <https://doi.org/10.1007/s00267-020-01332-4>
- USGS. (2021). Occurrence and Distribution of PFAS in Watersheds Across the United States. U.S. Geological Survey Scientific Investigations Report 2021-5107. <https://pubs.usgs.gov/sir/2021/5107/sir20215107.pdf>
- USGS. (2022). Preliminary findings on PFAS and floodwater interactions. U.S. Geological Survey Open-File Report 2022-1112. <https://pubs.usgs.gov/publication/ofr20221112>
- Wang, Z., DeWitt, J. C., Higgins, C. P., & Cousins, I. T. (2017). A never-ending story of per- and polyfluoroalkyl substances (PFASs)? *Environmental Science & Technology*, 51(5), 2508–2518. <https://doi.org/10.1021/acs.est.6b04806>
- Wang, Y., Chen, W., & Lu, H. (2025). Impact of urban flooding on PFAS contamination in surface waters. *Water*, 17(20), 2982. <https://doi.org/10.3390/w17202982>
- Wiedinmyer, C., et al. (2021). Impacts of aging sewer systems on chemical contaminant mobilization during flood events. *Journal of Environmental Engineering*, 147(4), 04021007. [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0001906](https://doi.org/10.1061/(ASCE)EE.1943-7870.0001906)
- Yao, Y., & Zhang, W. (2023). Socioeconomic factors and water contamination: Environmental justice in flood-prone communities. *Environmental Science & Policy*, 145, 324–331. <https://doi.org/10.1016/j.envsci.2023.01.013>
- Zhang, H., & Sun, X. (2024). PFAS transport dynamics in flood-affected urban watersheds. *Environmental Science & Technology*, 58(4), 2567–2576. <https://doi.org/10.1021/acs.est.3c07410>

Ethical, Legal, and Socio-Political Implications of AI By Owen Lewis^{1*}

Abstract

This paper integrates multiple perspectives on the legal, ethical, and creative challenges posed by artificial intelligence (AI), uniting discussions on governance, accountability, and human authorship. It explores how ethical concerns in AI are being addressed through organizational strategies, formal guidance, and stakeholder involvement, including the use of corporate social responsibility frameworks and participatory design. It further examines how existing legal frameworks, particularly tort law, handle liability and responsibility for harms caused by AI—highlighting complexities such as model drift and autonomous actions beyond human control. Real-world cases such as *People v. Wakefield* demonstrate the judicial challenges of AI-generated evidence, while *Britannica & Merriam-Webster v. Perplexity AI* illustrates tensions between innovation and intellectual property law. The discussion extends to the emerging debate over granting AI legal rights and protections, the evolution of AI in legal reasoning, and the struggle to reflect moral and contextual understanding within computational systems. Finally, it addresses the “human spark” requirement for authorship in copyright law and its implications for creative fields like choreography and comedy. Together, these sections reveal an evolving legal landscape that must balance innovation, fairness, and accountability in an increasingly AI-driven society.

Keywords Artificial intelligence; Legal reasoning;

I. Introduction

I structured this paper to guide the reader from the basic legal and ethical problems caused by AI to the bigger questions about rights, creativity, and how society should manage these technologies. It begins by examining liability, personhood debates, and the evolution of AI in legal reasoning. The paper concludes by addressing ethical governance and stakeholder responsibilities, showing how law, policies and society must adapt to an increasingly AI-driven world.

II. Legal and Ethical Challenges

Artificial Intelligence (AI) is changing many aspects of society, which is causing legal and ethical challenges. As AI systems become more a part of daily life, lawmakers face new questions regarding accountability, responsibility, and rights.

2.1. Liability and Responsibility for AI Caused Harms

2.1.1 Application of Traditional Tort Principles to AI-Related Harms

¹ Gretchen Whitney High School, Cerritos, California, United States

* Corresponding author

E-mail: owen_lewis@myabcusd.org (OL)

One of the first lessons is that current tort principles are being used to address immediate harms caused by AI tools. Courts mostly focus on human actions surrounding AI such as algorithm design and use, and then they assign liability accordingly. For example, if a poorly programmed AI causes damage, the deployer or programmer may be held liable under the duty of care [1]. Vicarious liability also holds organizations responsible when humans instruct AI tools in a harmful way.

2.1.2 Challenges of Model Drift and Ultra Vires AI Actions

However, as AI develops model drift, which is where AI behavior evolves beyond its initial programming without human intervention, then liability becomes more complex. Courts may compare such harms to known risks and hold humans accountable based on negligence principles [1]. But future cases might arise in which no clear human is at fault, especially with ultra vires actions which is when AI acts beyond its authorized scope [1]. This challenges traditional legal frameworks and requires innovative judicial reasoning.

2.1.3 Complications from Distributed AI and Proposed Legislative Responses

Furthermore, the distribution of AI over multiple systems further complicates accountability. When an AI spreads like a virus and causes harm such as defamation, tracing responsibility to a human or organization is close to impossible. Legislative measures like mandatory registration and insurance pools have been proposed to address this [1]. However, these solutions carry risks, including promoting reckless behavior.

These theoretical discussions of liability gain deeper significance when examined through real-world litigation. The following case, *People v. Wakefield*, illustrates how courts are already grappling with AI-based evidence and the balance between innovation, transparency, and due process.

Case Study: *People v. Wakefield*

i. Plaintiff

The prosecution argued that DNA evidence generated by TrueAllele was admissible under the Frye standard because the methodology is generally accepted in the scientific community. Validation studies, peer-reviewed research, and agency endorsements—including from NIST—supported its reliability. Although TrueAllele’s developer, Dr. Perlin, authored many studies, this is common in emerging technologies [2]. Courts in multiple states have admitted similar evidence without requiring source code disclosure, emphasizing that the Confrontation Clause does not guarantee access to proprietary algorithms.

ii. Defense

The defense contended that admitting TrueAllele violated both Frye and the Sixth Amendment of the US Constitution. Most validation studies were authored by Dr. Perlin himself,

raising bias concerns [2]. The inability to access the source code denied the defense a meaningful chance to assess reliability. The system's "black box" nature undermined due process, as the defendant could not confront the algorithmic evidence [3].

People v. Wakefield demonstrates the judiciary's early struggle with algorithmic opacity and accountability. As courts increasingly rely on AI-generated evidence, they must balance defendants' rights with proprietary protections, refining doctrines of admissibility for the algorithmic age.

2.2 Ethical and Legal Recognition of AI

2.2.1 Debate Over Granting Legal Protections and Rights to AI

Other than liability, a new question is rising about if AI should ever be given legal protections and rights? While some reject AI sentience outright, others think that AI might be smart enough to warrant protections [1]. The concept of personhood is crucial here. In the past "person" in law has also included corporations and natural resources [1]. Therefore extending certain rights to AI is legally possible.

2.2.2 Challenges and Opportunities in Extending Constitutional Protections to AI

The Fourteenth Amendment which says that laws must treat all people equally is a potential way for AI to get equal rights. However, it faces significant problems, such as standing and the non-human nature of AI [1]. Possible workarounds like guardianship might allow AI to bring legal challenges indirectly [1]. Furthermore, recent Supreme Court decisions signal a cautious and restrictive approach to expanding protections under this amendment [1].

2.3. Judicial Adaptability and Future Challenges

2.3.1 Evolving Judicial Approaches to Emerging AI Issues

An important lesson is that courts must adjust how they think to deal with new AI challenges. Traditional tort principles are effective for now, but new AI will test their limits. The judiciary will encounter cases involving autonomous AI that could act without direct human control [1]. Judges will also confront ethical debates about AI rights which are issues that have never been seen before in legal history [1].

2.3.2 Broader Societal and Policy Implications of AI

AI's impact influences not just legal issues but also society and public policy. A responsible AI design requires understanding both risks and benefits, which requires collaboration among technologists, lawyers, and policymakers [4]. Transparency, fairness, and accountability remain foundational principles in mitigating AI harms [4].

III. Evolution and Challenges of Artificial Intelligence in Legal Reasoning

AI has been expected for a long time to change how courts and legal systems understand and use laws. Since the early 1980s, researchers have tried to create computer models that mimic how legal reasoning works. This effort has involved developing systems that simulate how lawyers argue cases, retrieve legal information, and interpret statutory language. The main challenge is that it is hard to translate the inherently interpretive nature of law into precise rules or formulas that computers can use meaningfully [5-6].

3.1. Evolution of AI in Legal Reasoning

3.1.1 Rule-Based and Case-Based Reasoning

The early development of legal AI focused on rule-based reasoning (RBR), which uses logical if-then rules to decide legal cases, and case-based reasoning (CBR), which uses analogies to prior decisions. At first, these approaches were seen as competing; however, many researchers came to view them as complementary. Hybrid models that integrate both methods, such as CAsE-BAsed Reasoning Tool (CABARET) and Generator of Recursive Exemplar-Based Explanations (GREBE), demonstrated the potential of using cases to fill gaps left by incomplete rule sets [5].

3.1.2 Pioneering Systems and Models

The HYPO system, developed in the 1980s, was a significant program that demonstrated how computers could model legal arguments by analyzing appellate cases. It introduced the idea of "dimensions," which compares relevant features of cases. Later systems, such as CATO, built on HYPO to help law students learn how to make legal arguments. GREBE used analogical reasoning to reuse parts of past arguments in new cases, while CABARET combined rules and case comparisons to interpret tax laws [5].

3.1.3 Community Development and Scholarly Impact

The formation of the International Conference on AI and Law (ICAIL) in 1987 and the founding of the Artificial Intelligence and Law journal in 1992 showed that researchers from around the world were coming together to study how AI could be applied to legal matters. Researchers in Europe, Japan, and Australia contributed significantly to the field, exploring topics like ontologies, legal information retrieval, and defeasible reasoning [5].

3.1.4 Jurisprudence and Ethical Challenges

Despite advances, the field continues to struggle with the vague terms of legal language that don't have a precise definition. Scholars such as Thorne McCarty have argued that AI should engage more with jurisprudence, particularly the challenge of balancing strict legal rules with the flexible judgement that judges use [5]. The inability to fully capture the moral and contextual aspects of legal reasoning remains a central limitation of AI models [5-6].

IV. Evolution of AI and Its Legal Implications

The rise of AI has posed major questions for legal systems that were built around human creativity. In U.S. copyright law, authorship belongs only to humans, but now AI can make art, writing, and music mostly on its own, which challenges how the current law works. [7]. Courts and agencies, especially the U.S. Copyright Office have stated that a person must be involved to claim ownership. However, as AI grows more advanced, distinguishing between human and machine creations is becoming more difficult, forcing artists and legal experts to rethink what counts as creative labor [8].

To see these challenges in action, the following contemporary dispute—Britannica and Merriam-Webster v. Perplexity AI—illustrates how AI innovation tests copyright boundaries.

4.1 Case Study: Britannica and Merriam-Webster v. Perplexity AI

i. Plaintiff

Britannica and Merriam-Webster allege that Perplexity AI unlawfully copied and repurposed their copyrighted materials, using retrieval-augmented generation (RAG) to produce summaries and outputs that divert readers and revenue[9]. Plaintiffs argue that Perplexity's actions are non-transformative, purely commercial, and substantially harmful to their markets, even amounting to trademark dilution through false attributions.

ii. Defense

Perplexity asserts fair use, claiming its RAG model is transformative and functional—using excerpts to answer questions, not replicate articles. The platform cites sources, links to originals, and discloses AI generation. The company positions itself as an innovator, not an infringer, warning that an adverse ruling could stifle technological progress [9].

This dispute captures a pivotal moment in copyright law. Courts must determine whether AI's derivative use of protected material constitutes transformation or infringement. As with *People v. Wakefield*, transparency, intent, and accountability remain central to legal analysis.

4.2 The “Human Spark” Requirement in Copyright Law

Modern copyright law insists that only works created by humans are eligible for protection, a principle reinforced by courts and the U.S. Copyright Office in recent cases involving AI-generated content [7]. The key rule is whether a work contains a human "spark" which is a clear sign of human creativity that shapes the final product. Just pressing a button or generating content through prompts is not enough, the person has to add real creative effort for the work to be protected by copyright [8].

4.2.1 Choreography: A Case Study in Human Creativity

Choreography demonstrates how AI can be used as a tool without replacing the creator. For example Sir Wayne McGregor uses AI to generate sequences based on his previous work,

however he is the one who curates, refines, and teaches the movements to the dancers [8]. Legally, choreography must be created for humans to perform to qualify for copyright, which makes sure that no matter how advanced AI becomes, it cannot be the sole creator of a dance [8].

4.2.2 Comedy: The Implicit Requirement for Humanity

In comedy, while written jokes and stand-up routines can be copyrighted, both audiences and performers want a level of human authenticity that AI cannot replicate [8]. Comedians who have experimented with AI-generated material, such as Anesti Danelis and Karen Hobbs, found that while the technology can offer ideas, things like timing, delivery, and audience connection still requires a person. Due to comedy needing people to perform well, this helps protect the art form being fully taken over by AI [8].

4.2.3 Legal Uncertainty and the Use of Private Law

Both choreography and comedy suffer from unclear copyright protections, especially for things like live performances, improvisations, and how they are delivered, because these are hard to record or write down [6]. Due to this, people in these fields use private agreements, licenses, and community rules to protect their work. These systems allow creators to control how their work is shared, performed, or reproduced, even when copyright law does not offer strong protection [8].

V. Ethical Governance of Artificial Intelligence

AI is transforming society, but this progress brings ethical risks such as bias, lack of transparency, and unfair impacts. These concerns require more than technical fixes. Organisations, policymakers, and citizens must help guide how AI is developed and used [10]. Addressing AI ethics involves layers of effort. Companies adopt ethical policies, governments create laws and standards, and civil society advocates for rights and justice. This paper examines these efforts, focusing on organisational approaches, ethical guidance tools, and stakeholder roles.

5.1 . Organisational Strategies for AI Ethics

5.1.1 Governance and Risk Management

Organisations are adopting tools to manage AI risks. These include ethics committees, impact assessments, and bias detection systems. These tools help align AI use with social values—and can reduce legal risk if harm occurs [10]. Good governance often combines technical fixes with human oversight, training, and partnerships. These efforts aim to increase accountability while keeping AI systems adaptable to changing conditions.

5.1.2 Strategic Commitments and CSR

Many companies are framing AI ethics within corporate social responsibility (CSR). They work to understand how AI affects customers, employees, and society at large [10]. Some adopt human rights frameworks to guide their development practices. Engaging stakeholders—especially those affected by AI—is central to CSR. It helps companies identify ethical risks early and respond to public expectations more effectively.

5.1.3 Insights from Real-World Practice

Research shows that while organisations care about AI ethics, they often focus on visible issues like data privacy or bias. Broader concerns, such as long-term social impacts or human rights, are less addressed [10]. This gap suggests a need for more robust and consistent adoption of ethics strategies. Many tools exist, but they are not evenly applied across sectors.

5.2. Guidance Mechanisms in AI Ethics

5.2.1 Ethics Guidelines and Frameworks

Many governments and organisations have issued ethics guidelines. The EU’s “Trustworthy AI” framework is a well known example [10]. These frameworks promote principles like fairness, transparency, and accountability. However, these guidelines can be vague or hard to apply. Without clear enforcement, they risk becoming symbolic rather than effective [10].

5.2.2 Standards and Professional Codes

Standards bodies like IEEE and ACM have updated their ethical codes for AI. IEEE’s P7000 series addresses issues such as algorithmic bias and transparency [10]. These standards aim to make ethics more concrete. Certification schemes may eventually help enforce them in high-risk AI systems.

5.2.3 Design Methods for Ethical AI

Design methods like “ethics by design” and “value-sensitive design” embed ethical thinking into development processes. They help teams reflect on long-term impacts early in the design phase [10]. These methods address the “Collingridge dilemma,” the challenge of anticipating harm before technology becomes hard to change [10].

5.2.4 Tools and Templates

Many ethics tools support teams during AI development. These include checklists, risk templates, audit frameworks, and stakeholder maps. Examples come from groups like AI Now Institute, Doteveryone, and PwC [10]. While not perfect, such tools are growing in popularity. They help translate high-level principles into real-world action.

5.3. Stakeholders in AI Ethics

5.3.1 Policymakers and Public Bodies

Policymakers shape AI ethics through laws, funding, and standards. Key actors include the EU, UN agencies like UNESCO, and national ethics boards [10]. Laws and regulations are often enforced at the national level, while global bodies influence best practices. Public funding also steers AI research priorities.

5.3.2 Organisational Stakeholders

This group includes tech companies, insurers, schools, media, and professional bodies. Some build AI systems; others help shape public understanding or create standards [10]. Media coverage of AI ethics has pressured companies to act. Educational institutions are also raising awareness and training future developers in ethical practices.

5.3.3 Individuals and Affected Groups

Individuals play many roles. Developers, users, activists, and even hobbyists all influence AI. Some people, however, are affected by AI without having a voice. Examples include prisoners judged by risk algorithms or patients diagnosed by AI tools. These silent stakeholders are often overlooked, even though they may suffer most from AI's harms [10].

VI. Conclusion

Artificial intelligence is transforming both society and the legal landscape, presenting unprecedented challenges that strain traditional legal and ethical frameworks. While existing tort law principles provide a useful starting point for addressing AI-caused harms, emerging issues such as autonomous decision-making, distributed AI systems, and unclear lines of liability demand innovative legal approaches and legislative reform. The debate over granting AI legal rights further complicates the picture, raising profound questions about personhood, accountability, and justice. In the realm of legal reasoning and copyright law, AI's growing capabilities challenge long-held notions of creativity, authorship, and human uniqueness. As AI continues to evolve, it is imperative for courts, lawmakers, technologists, and artists to collaborate in crafting adaptive, transparent, and ethical responses. Ensuring AI systems are not only effective but also fair, accountable, and respectful of human values will be crucial to safeguarding legal integrity and public trust in the digital age.

Works Cited

- [1]- “The Ethics and Challenges of Legal Personhood for AI.” The Yale Law Journal, 22 April 2024,
<https://www.yalelawjournal.org/forum/the-ethics-and-challenges-of-legal-personhood-for-ai>. Accessed 16 August 2025.
- [2]- People v. Wakefield. New York Court of Appeals, No. 3, decided 26 Apr. 2022. Justia,
<https://law.justia.com/cases/new-york/court-of-appeals/2022/3.html>.
- [3]- New York Supreme Court, Docket No. 2014099, decided. FindLaw,
<https://caselaw.findlaw.com/court/ny-supreme-court/2014099.html>.
- [4]- Gordon, JS. AI and law: ethical, legal, and socio-political implications. *AI & Soc* 36, 403–404 (2021).
- [5]- Prakken, Henry. *AI & Law, Logic and Argument Schemes*. 2006. University of Groningen, final author’s version. PDF file.
- [6]- Loui, Ronald. "AI and Law: A Fruitful Synergy." *Artificial Intelligence*, vol. 150, no. 1, 2003, pp. 1–20. Elsevier
- [7]- Legal Analysis on the U.S. Copyright Office’s Human Authorship Requirement and Judicial Responses to AI-Generated Works.
- [8]- The Human Spark: Copyright, Authorship, and the Boundaries of AI, *Harvard Law Review*, Vol. 137, 2024.
- [9]- *Encyclopædia Britannica, Inc. and Merriam-Webster, Inc. v. Perplexity AI, Inc.* Complaint, No. 1:25-cv-7546, U.S. Dist. Ct., S.D.N.Y. Filed Sept. 10, 2025. Susman Godfrey,
<https://www.susmangodfrey.com/wp-content/uploads/2025/09/Britannica-v.-Perplexity-Complaint.pdf>
- [10]- Stahl, B.C. (2021). Addressing Ethical Issues in AI. In: *Artificial Intelligence for a Better Future. SpringerBriefs in Research and Innovation Governance*. Springer, Cham.
https://doi.org/10.1007/978-3-030-69978-9_5

AI and The Environment: Can Climate Policies Protect Our Planet? By Justin Y Kim

Abstract

Artificial Intelligence (AI) tools like ChatGPT use a remarkable amount of electricity and water behind the scenes. As these systems become more popular, their hidden environmental costs, especially energy use and water consumption, are a serious concern. This paper will explore how new laws and policies can help reduce the damage AI is causing to the planet through its energy and water consumption. The paper will ultimately evaluate different solutions to mitigate AI's environmental harm. AI is a powerful tool, but without immediate actions, AI could quietly become one of the world's biggest environmental threats.

Keywords

Artificial Intelligence, Environmental Sciences, Public Policy, Climate Change, Fossil Fuels, Water, Electricity, Green AI. `

Introduction

Laws and policies maintain order in our society, and while their contributions are crucial in maintaining harmony, our focus has shifted away from the environment. As our world evolves, new generations are adopting new tools such as Artificial Intelligence (AI), which have significantly impacted our society. AI has expanded into one of the greatest inventions mankind has seen. Yet, the benefits of AI have slowly manipulated the minds of our society, as it perfectly disguises itself from the harm it secretly does.

AI has been around for nearly seven decades, originally used in game-playing machines and problem-solving for simple tasks. Over the past several years, our world has seen some of the most significant developments in AI. Now, such AI exists that can understand and respond to human conversations; these are called Large Language Models (LLM). Despite the development of such an extraordinary tool in such a short amount of time, behind the scenes, the rapid development of AI is slowly decaying our environment. Since each individual AI needs extremely powerful computing power, each question input into the algorithms needs significant energy consumption to function as rapidly and accurately as AI does. Additionally, extensive training is necessary to ensure the models incorporate the most recent and accurate data from billions of databases, thereby reducing hallucinations and bugs.¹ AI databases compute vast amounts of data every second, all year round. Large amounts of water are needed to cool down the operating systems that overheat, preventing the system from crashing.

Various approaches are needed to solve this issue.² The approach I will outline in this paper depicts the usage of policies to attempt to decrease the environmental harm that AI causes without slowing down its development. How can new policies help mitigate the environmental damage caused by AI's energy and water consumption? Our planet is the most important aspect of living. Simply, without our planet, we would not be here. Yet, with such rapid development of AI models, our planet is rapidly degrading as well. Although the impacts are not immediately

noticeable, without solving these issues, they can grow to be unstoppable in the foreseeable future, and such issues include frequent extreme weather changes, rising sea levels, and disruptions in ecosystems.

Literature Review

1. AI is double-dealing the environment.

The fast-paced development of Artificial Intelligence, particularly LLMs (large language models), is ultimately resulting in the neglect of the environment. An analysis illustrates the magnitude of the impact pollution has and will continue to have on our environment.³ Some examples include Sea Levels in America 1 by 6.6 feet in 2100, Global Temperatures rising (i.e, the Summer of 2023 was the warmest summer ever recorded, it was on average 2.1 degrees Fahrenheit more than the summer of 1980). If the ice sheets in the Arctic continue to decrease at the current pace, the once thriving Arctic Ocean will be ice-free before mid-century.

1.1 Water consumption

The rapid pace of AI development is affecting our environment in three specific ways. Firstly, the new AI systems need to undergo extensive training and coding. Once the AI system is ready, it receives billions of prompts in a single day. This results in the AI systems overheating. Most well-known AI companies use water to cool down their systems because it is very cost-effective. A lot of water is necessary for cooling down the hardware used for designing, building, and deploying AI models, ultimately, this results in multiple strains on water sources, ecosystems, and water supplies.⁴ Companies resort to using water to cool down their systems because it is more cost and time-efficient than using air cooling mechanisms. Water can absorb heat more efficiently than air, which allows the AI systems to work at full power without overheating. While companies may be saving money by not constructing more AI processors, they are also harming the environment by consuming water from diverse ecosystems.⁵ Overusing water damages environmental flows, destroys wetlands, diverse wildlife, and the biodiversity of the landscape. Even though developing new fast and efficient AI processors that can function at a rapid pace is important, it is also important to realize how changes can be made to decrease the water intake, because overusing resources can have irreversible damage.

1.2 Electricity

The second reason is that AI consumes a significant amount of electricity, primarily sourced from fossil fuels. For every task AI is prompted on, specifically LLM models, the chatbot must review billions of databases to find the most relevant and accurate information in mere seconds. This extensive computing ultimately results in the machines in the data centers consuming massive amounts of electricity.⁶ Artificial Intelligence uses massive quantities of fossil fuel-based electricity, which largely impacts greenhouse gas emissions. Not only do AI facilities consume excessive energy during operation, but the process of training and designing

models also requires a significant amount of energy and results in the release of numerous greenhouse gases.⁷ The International Energy Agency (IEA) estimates electricity used by Artificial Intelligence facilities in 2022 was 240-340 TWh, in other words, nearly 1.3% of the total energy demand in the world. Popular Artificial Intelligence chatbots like ChatGPT and Google Gemini have only been around for several years. Despite their recent development, they have already grown to take up more than 1 percent of the world's energy consumption.

Although the consequences of this aren't as noticeable right now, they will become more apparent in the next 5-10 years.⁸ By 2026, electricity consumption from artificial intelligence could reach 4% of annual global energy usage — similar to the electricity usage in the entire country of Japan.⁹ The amount of energy used by newly developed AI models could be extremely detrimental to the environment as they will use more energy than can be created. It is not just the energy that AI consumes; AI also consumes dangerous amounts of fossil fuels. AI uses massive quantities of energy, and most of that energy comes from the burning of fossil fuels. The Earth has a certain amount of fossil fuels that will eventually run out. Once all of Earth's fossil fuels are depleted, chaos will strike as most of the energy runs on fossil fuels. Thus, preventing or delaying this from happening is necessary.

2. Green AI

It is crucial to remember that while mitigating the harm caused by AI is important, slowing down its development is not an acceptable solution. As explained in sections 1.1 and 1.2, AI is harming our environment at an alarming rate. In just three years of development, Artificial Intelligence has already consumed the same amount of energy as the entire country of Japan. Thus, previous solutions have been attempted, one of the major attempts being the attempt to spread the usage of Green AI.

2.1 Benefits of Green AI

Green AI represents specific models of AI that aim to decrease environmental harm. Green Artificial Intelligence decreases carbonization.¹⁰ Additionally, Green Artificial Intelligence focuses on sustainable and efficient AI Models.¹¹ Not only this, but Green Artificial Intelligence aims to create algorithms that aim to create energy-efficient Artificial Intelligence models and explore eco-friendly approaches to AI while enhancing hardware optimization. Although Green AI is relatively new, it has already demonstrated its potential to help the environment by reducing its carbon footprint and developing AI systems that overheat less than major companies like Google, Gemini, and ChatGPT.

2.2 Negatives of Green AI

Although Green AI is an effective solution for reducing the harm caused by current Artificial Intelligence models, major issues were not previously explored in Section 1.1. The first issue is the cost. Artificial Intelligence already costs a lot of money, the headlines of a CNBC news report, “Tech Megacaps plan to spend more than \$300 billion in 2025 as the Artificial

Intelligence race intensifies".¹² This means that keeping AI systems running worldwide costs 300 billion yearly, and with the increase of widespread use of technology, this cost is expected to grow exponentially over the course of our future. Furthering this point, developing new AI models that can reduce carbonization and overheating rates while maintaining current functionality would require a significant investment. Sousa 25' furthers this point by including that developing Green Artificial Intelligence is very resource-intensive. This would result in major companies like Google, Gemini, and Chat GPT re-routing their funding to Green AI efforts instead of developing and technologically advancing these current widespread AI systems. However, ceasing and slowing down the advancements of Artificial Intelligence is extremely harmful. Artificial Intelligence needs to develop at a rapid pace, particularly in addressing space-related issues, such as the potential consequences of Earth's demise. If that were to happen, having advanced AI models to help space exploration specifically is essential for the future of humankind.

Additionally, another issue with implementing Green Artificial Intelligence as a mainstream AI is the quality of its models. A challenge of Green Artificial Intelligence is the quality and availability of data. Consider that Google Gemini, along with other generative AI models, has been refining its models for many years.¹³ Starting again from scratch with Green AI would also be time-consuming.

3. Policies in Artificial Intelligence

Artificial Intelligence is the future of our planet; Generative AI needs to develop and advance technologically while still maintaining the environment. Thus, developing new systems from the start will be time-consuming. Instead, implementing policies for large AI companies could be a solution. Some companies already publicly announce their environmental impacts, for example, Amazon and Microsoft. Despite this, no current laws or policies mandate these companies to tell the truth. This is the reason why policies need to be implemented to mandate companies to reveal their true impacts on the environment. However, AI is utilized by millions of people, and it is not just the companies that need to change. By large companies that are responsible for large environmental impacts, revealing their environmental impacts might educate the general public on the harm AI does to the environment. Rules can help people in a supportive way. When companies release data about their environmental impact, such as water, fossil fuels, and electricity consumption, it can also affect the public's view of AI.

3.1 Benefits of Implementing Policies in AI

Well-known companies like Amazon and Microsoft have already published annual reports on their climate-related issues and commitments. Yet, no laws regulate their reports to ensure the accuracy of their statements. Once government policies are created, companies that previously fabricated reports or never submitted one will be forced to do so. The phrase "numerical framing" is used quite frequently throughout his article, *The Psychology of Numbers*. In this article, the authors weigh down the fact that numbers can change perceptions.¹⁴ This is

key to swaying the minds of the general public. Once they see the actual numbers on how much harm AI is doing to our environment, they will realize that every unnecessary prompt that they enter into Generative AI or another type of AI has devastating harm to our environment.

Eighty percent of people globally want more decisive climate action from their government.¹⁵ Public involvement in environmental governance is crucial to maintaining a healthy relationship with human civilization and our Earth.¹⁵⁻²⁷ By implementing policies that mandate companies to release annual reports on their environmental impacts, the public will be motivated to take action, as they will realize the government supports further climate action. Finally, policies are significantly more cost-effective than designing Green AI, as creating new systems and generators would be more expensive than developing a policy.

3.2 Negatives of Implementing Policies in AI

There are three major issues that arise when policies are implemented in AI. The first issue is who will enforce the policy? Even if the policy is implemented in AI companies globally, who will be the international organization that investigates each company to ensure they're actually following the policy? An article written in 2022 emphasizes that "An unenforced law is worse than no law at all".²⁸ The article highlights the value that people will not follow a rule unless there is a repercussion for breaking that rule. Ultimately, without an international force that will investigate and criminally charge companies that do not follow the policy, the policy is useless. With an international force comes more problems. Who will fund the organization? How will they enforce the policy? Would they have to create a whole new court system?

The second issue in implementing AI policies is determining what they will cover. AI spans such a broad range of intelligence that it is nearly impossible to include a one-size-fits-all rule. The third major issue is when the laws will come into effect. As stated before, it is impossible to invoke a one-size-fits-all rule when it comes to AI. Policies take a long time to pass, with multiple policies needing to be passed, it would take an extraordinary amount of time for each policy to be passed. It is important not to forget that artificial intelligence is constantly evolving and developing as we speak. It will be impossible for the policy to keep pace with the constant evolution of artificial intelligence.

Discussion and Conclusion

AI is a fundamental tool for the future of human life in this universe; however, to maintain our own planet we live on, we must decide whether developing policies to regulate the environmental harm artificial intelligence causes or converting AI to Green AI is the better solution. AI imposes a multitude of impacts on our environment. AI systems require a large amount of electricity and water to function properly without overheating. Secondly, for AI systems to function properly, they require a massive amount of electricity. However, the electricity is generated from fossil fuels. Ultimately, this means that AI is rapidly consuming two of Earth's important dwindling resources. Green AI could be a plausible solution to the environmental impacts caused by AI. Creating a completely new form of AI could not only

enhance the technological advancements of AI while also supporting climate change mitigation. However, using Green AI could cause setbacks in future development and be financially devastating for AI companies. Another plausible solution to limit the climate damage caused by AI could be to implement policies requiring companies to disclose their environmental impacts. In fact, evidence suggests that people are willing to make lifestyle changes in order to improve the climate. However, a major flaw in this solution is the time. The amount of time it takes to pass a policy would not keep pace with the rapid development of AI.

Works Cited

1. Zimmerman, A. Not a Blank Slate: The Role of Big Tech in Misinformation and Radicalization. *Digit. Soc.* 2024, 3 (6), 91. <https://doi.org/10.1007/s44206-024-00091-5>.
2. Chapman, M.; Xu, L.; Lapeyrolerie, M.; Boettiger, C. Bridging Adaptive Management and Reinforcement Learning for More Robust Decisions. *Philos. Trans. R. Soc., B* 2023, 378 (1881), 20220195. <https://doi.org/10.1098/rstb.2022.0195>.
3. NASA. The Effects of Climate Change. *NASA Sci.* 2024, October 23. <https://science.nasa.gov/climate-change/effects/>.
4. Zewe, A. Study: AI Could Lead to Inconsistent Outcomes in Home Surveillance. *MIT News* 2024, September 19. <https://news.mit.edu/2024/study-ai-inconsistent-outcomes-home-surveillance-0919>.
5. Brabeck-Letmathe, P. The Damage Caused by Water Overuse. *World Econ. Forum* 2014, December 10. <https://www.weforum.org/stories/2014/12/the-damage-caused-by-water-overuse/>.
6. Kandemir, M. Why AI Uses So Much Energy—And What We Can Do About It. *Penn State Inst. Energy Environ.* 2025, September 8. <https://iee.psu.edu/news/blog/why-ai-uses-so-much-energy-and-what-we-can-do-about-it>.
7. Chen, S. How Much Energy Will AI Really Consume? The Good, the Bad and the Unknown. *Nature* 2025, 639 (8053), 22–24. <https://doi.org/10.1038/d41586-025-00616-z>.
8. Yale School of the Environment. Can We Mitigate AI’s Environmental Impacts? *Yale Sch. Environ.* 2024, October 10. <https://environment.yale.edu/news/article/can-we-mitigate-ais-environmental-impacts>.
9. Trehan, A. Bias in Green AI: Addressing Disparities in Data and Algorithms. In Williams, P.; Kulkarni, S. (Eds.) *Advancing Social Equity through Accessible Green Innovation*; IGI Global: Hershey, PA, 2025; pp 63–76. <https://doi.org/10.4018/979-8-3693-9471-7.ch005>.
10. Bolón-Canedo, V.; Morán-Fernández, L.; Cancela, B.; Alonso-Betanzos, A. A Review of Green Artificial Intelligence: Towards a More Sustainable Future. *Neurocomputing* 2024, 599, 128096. <https://doi.org/10.1016/j.neucom.2024.128096>.
11. Subin, S. Tech Megacaps Plan to Spend More Than \$300 Billion in 2025 as AI Race Intensifies. *CNBC* 2025, February 8. <https://www.cnbc.com/2025/02/08/tech-megacaps-to-spend-more-than-300-billion-in-2025-to-win-in-ai.html>.
12. UST. What Is Green AI? UST Explainers n.d. <https://www.ust.com/en/ust-explainers/what-is-green-ai>.
13. El Mahyaoui, N. Numbers Are Everywhere, Guiding Our Choices, Shaping Perceptions, and Influencing Emotions. *LinkedIn* n.d. Retrieved August 26, 2025, from <https://www.linkedin.com/pulse/psychology-numbers-nabil-el-mahyaoui-ytzwe/>.
14. University of Oxford. 80 Percent of People Globally Want Stronger Climate Action by Governments According to New Survey. *Univ. Oxford* 2024, June 20.

- <https://www.ox.ac.uk/news/2024-06-20-80-percent-people-globally-want-stronger-climate-action-governments-according-new>.
15. Dai, M. How to Enhance Public Participation in Environmental Governance? Evidence from China. *Sustainability* 2024, 16 (8), 3154. <https://doi.org/10.3390/su16083154>.
 16. Jodoin, S.; Duyck, S.; Lofts, K. A. Public Participation and Climate Governance: An Introduction. *Rev. Eur. Comp. Int. Environ. Law* 2015, 24 (2), 117–122. <https://doi.org/10.1111/reel.12126>.
 17. Overney, C. Designing for Constructive Civic Communication: A Framework for Human–AI Collaboration in Community Engagement Processes. *arXiv* 2025, arXiv:2505.11684. <https://arxiv.org/abs/2505.11684>.
 18. Fontes, C.; Al Haddad, C.; Antoniou, C.; Lütge, C. Assessing the Collective Challenges of AI: An Applied Framework Focused on People Centeredness and Public Participation. *Urban Gov.* 2024, 4, 245–254. <https://doi.org/10.1016/j.ugj.2024.12.003>.
 19. Wu, J.; Yu, Z.; Wang, R. The Impact of Public Participatory Environmental Regulation on Carbon Emission Intensity: A Policy Text Analysis. *Front. Environ. Sci.* 2025, 13, 1534066. <https://doi.org/10.3389/fenvs.2025.1534066>.
 20. Mocanu, M.; Bibiri, A.-D.; Rusu, V. D.; Moroşanu, A.; Bejan, I. G. Enhancing Civic Engagement with Science: A Comparative Approach across European Regions. *Scientometrics* 2025, 130 (1), 447–468. <https://doi.org/10.1007/s11192-024-05198-7>.
 21. Sánchez, F.; Olmos, R.; Sandoval, L. A.; Casani, F. Psychosocial Factors That Favor Citizen Participation in the Generation of Scientific Knowledge. *Scientometrics* 2024, 129 (8), 5021–5036. <https://doi.org/10.1007/s11192-024-05099-9>.
 22. Züger, T.; Asghari, H. AI for the Public: How Public Interest Theory Shifts the Discourse on AI. *AI Soc.* 2023, 38 (2), 815–828. <https://doi.org/10.1007/s00146-022-01480-5>.
 23. Zehner, N.; Ullrich, A. Dreaming of AI: Environmental Sustainability and the Promise of Participation. *AI Soc.* 2025, 40 (4), 2605–2617. <https://doi.org/10.1007/s00146-024-02011-0>.
 24. Matveieva, O.; Navumau, V.; Gustafsson, M. Adoption of Public E-Services versus Civic Tech Services: On the Issue of Trust and Citizen Participation in Ukraine and Belarus. In Janssen, M.; Csáki, C.; Lindgren, I.; Loukis, E.; Melin, U.; Pereira, G. V.; Bolívar, M. P. R.; Tambouris, E. (Eds.) *Electronic Government: 21st IFIP WG 8.5 Int. Conf., EGOV 2022, Linköping, Sweden, Sept 6–8, 2022, Proc.*; Springer: Cham, Switzerland, 2022; pp 15–30. https://doi.org/10.1007/978-3-031-15086-9_2.
 25. Ferdman, A. AI, Deskillling, and the Prospects for Public Reason. *Minds Mach.* 2025, 35 (3), 38. <https://doi.org/10.1007/s11023-025-09737-w>.
 26. Díaz, P.; Aedo, I. Technology-Mediated Citizen Participation in Emergency Management: Balancing Social Capital and Efficacy. In Bramwell-Dicks, A.; Evans, A.; Winckler, M.; Petrie, H.; Abdelnour-Nocera, J. (Eds.) *Design for Equality and Justice. INTERACT 2023 Workshops*; Springer: Cham, Switzerland, 2024; pp 44–56. https://doi.org/10.1007/978-3-031-61698-3_4.

27. Hill, D.; Melander, A. Strategic Design for Public Purpose: Enriched and Expanded Practices for Human-Technology Interaction. In Malakhatka, E.; Wiberg, M. (Eds.) Human-Technology Interaction: Interdisciplinary Approaches and Perspectives; Springer: Cham, Switzerland, 2025; pp 23–80.
<https://link.springer.com/book/10.1007/978-3-031-78357-9>.
28. The Columbian. In Our View: An Unenforced Law Is Worse Than No Law at All. Columbian 2022, September 13.
<https://www.columbian.com/news/2022/sep/13/in-our-view-an-unenforced-law-is-worse-than-no-law-at-all/>.

Designing Classrooms to Accommodate Children with ADHD-related Sensory Perception Issues By Adhya Gupta

Introduction

ADHD or Attention deficit/hyperactivity disorder is a neurodevelopment condition characterised by inattention, hyperactivity and impulsivity. The symptoms of attention deficit hyperactivity disorder (ADHD) are relatively common. Current estimates place the prevalence of ADHD among Indian children at 7.1%, although there are significant variations based on region, gender, and socioeconomic status (Gore and Morgan).

Sensory processing refers to the neurological mechanism by which sensory information from the environment and the body is received, organized, and interpreted to produce adaptive behavioural response. Atypical sensory processing patterns (SPPs) can impact the development of behavioral, emotional, motor, and cognitive domains. While SPPs can occur independently, they are usually correlated with neurological or developmental disorders. Most research on SPPs and neurodivergence focuses on children and adults with autism spectrum disorder (ASD). However, SPPs also occur more frequently in people with ADHD than in neurotypical individuals. (Ghanizadeh; Jurek et al.)

At a functional level, children with ADHD may be overwhelmed by environmental stimuli such as lights or sounds. On the other hand, they may not recognize or respond to cues from teachers or peers. SPPs can thus make classroom environments particularly challenging for children with ADHD.

This paper examines some of the ways in which SPPs manifest in children with ADHD, and some proven measures that can be adopted by schools in India to accommodate the needs of children with ADHD.

How SPPs manifest in children with ADHD

Researchers have classified sensory inputs into eight categories (Lonkar):

Visual: Colours, light / dark, shapes and sizes

Auditory: Volume, pitch and rhythm of sounds

Gustatory: Taste and mouthfeel

Olfactory: Odours

Tactile: Textures, external temperature, touch,

Proprioception: Movement and muscular activity

Vestibular: Balance and position in space

Interoception: Sensations from internal organs

Our bodies are able to experience and respond to these seven kinds of stimuli. Sensitivity and response fall on a range from low to high (Butera et al.).

- Sensitivity to a stimulus: Hyposensitive / High threshold or Hypersensitive / Low threshold
- Self-regulation in response to that stimulus: Passive to Active

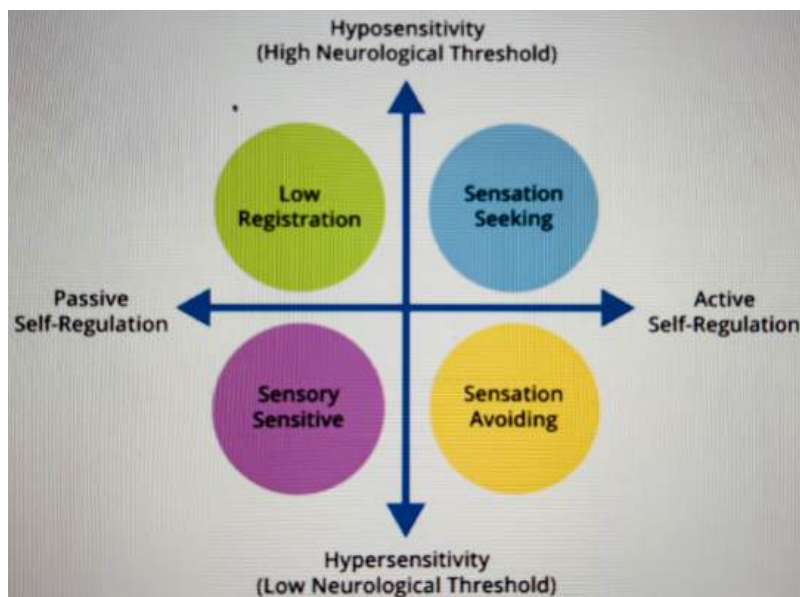


Fig 1: Four sensory modulation domains of SPPs based on sensitivity and self-regulation (ConnectAbility)

While there are no absolute scales of sensitivity and response, a normal range can be established based on the typical responses of individuals. SPPs that result in atypical sensory modulation can present in four distinct patterns shown in Figure 1.

Jurek et al. provide some examples of how these SPPs can be seen in children with ADHD:

Sensation seeking: The child may touch objects, play with stationery, drum their fingers on desks, or fidget in other ways to stay alert.

Low registration: The child does not respond immediately when called, and appears to be daydreaming or otherwise overly engrossed in some activity.

Sensory sensitive: The child is distracted by stimuli (lights, sounds) and cannot focus on the task at hand

Sensation avoiding: The child is agitated by stimuli and tries to distance themselves from the stimulus, or throws a tantrum until the stimulus stops.

Just as ADHD is a condition with heterogeneous presentations, not all children with ADHD will respond the same way to a stimulus. Ghazanideh notes that while sensory over-responsivity is more common among children with ADHD, this is mostly seen with regards to tactile, auditory, and visual stimuli. The same child may display differing SPPs to different sensory inputs, e.g. sensory sensitivity to auditory inputs but sensation seeking for tactile inputs.

SPP incidence among children with ADHD is significantly higher than neurotypical children, but some SPPs are more frequent than others. Low registration was 11% higher in children with ADHD, whereas sensory seeking was 69% higher (Jurek et al.)

Need for designing an ADHD friendly classroom

Children with ADHD behave differently in school, at play, and in social situations depending on their environment. In order to improve focus on a specific task, treating ADHD requires avoiding stimulants and distractions. According to Ghazali, the learning environment and its components have been crucial in enhancing classroom instruction and have a big impact on students' learning. Some of the challenges that children with ADHD frequently experience have a direct impact on their mental health. Parents, schools, or other careers must take a variety of precautions to regulate the disorder, as there is no known treatment. Therefore, creating an environment that helps manage a child's symptoms of ADHD is necessary to aid in their recovery (Ghazali et al.)

An educational environment must encourage the development of healthy skills, but for children with ADHD, this is not the case most of the time. These children often have poor grades, poor reading and math skills, and are more likely to repeat a grade (Daley and Birchwood).

Primary school is when many of these children are first diagnosed because this period of childhood is when they experience school and all the structure that comes with it for the first time (Holland and Sayal). Even with specific symptoms identified, some individuals are not diagnosed until much later in life. 35% to 65% of these individuals will continue to show symptoms into adulthood. On average, children are diagnosed around the age of 7, with a range of 5 years for severe cases and an age of 8 for mild cases. (Holland and Sayal). Severity is based on the type and number of symptoms shown. No matter what symptoms are prominent in childhood the hyperactive symptoms tend to decline as they move into adolescence, and inattentive symptoms tend to stay more stable. Even if this shift does occur and some symptoms disappear over time, as long as the specified number of symptoms are still present, the ADHD diagnosis remains.

Current estimates of the incidence of ADHD among Indian children are at 7.1%, making the overall number of children with ADHD in India at around 31 million (Gore and Morgan). Or to put it another way, on average there will be 2 or 3 children with ADHD in the typical Indian classroom with 40 students. It is therefore crucial that steps are taken to ensure that this large number of children experience a school environment that is mindful of their needs.

Designing an ADHD-friendly classroom

Accommodations for children with ADHD take many shapes. The most common of these are the provision of additional time to complete tests, exams, and assignments. Others include the provision of tools such as calculators. (Lovett and Nelson). However, these accommodations focus on testing outcomes and do not necessarily improve the learning experience. Rethinking

the classroom environment to account for SPPs could produce a more holistic improvement in learning for children with ADHD.

- Colours: Bright colours and high contrast environments are more likely to hold children's attention, which would help children with low visual registration.
- Visual Timers: Large, prominent clocks help students internalize the passage of time, reducing the "time blindness" often associated with the disorder. Pairing this with an audio alert may improve their efficacy in combating time blindness.
- Single-seat desks: Allowing each child to sit by themselves would help those with tactile SPPs, especially with regards to being in close proximity to other people.
- Sound insulation: Reducing intrusive noises from outside the classroom will aid children with auditory over-responsivity.
- Soft-touch materials: Classroom furniture made from soft-touch materials serve two purposes. First, it can provide low-grade tactile stimulation that is useful for sensation seeking children. Second, it will absorb sound and reduce auditory over stimulation.
- Get-away area: A designated space within the classroom, where a child can sit and self-soothe if they feel overwhelmed

In addition to the design of the classroom, creating an accommodating environment for children with ADHD involves removing ambiguity and ensuring that clear cues are given at all points. Some ways of ensuring clarity in the classroom include:

1. Clearly demarcated spaces: Where possible, spaces for quiet reading, spaces for group work, spaces for discussion should be demarcated in the classroom. Children will be cued into what kind of behaviour is expected of them in each space and are more likely to act in an appropriate way.
2. Clear signals at the start and end of a task: Teachers can build a routine around specific activities – reading silently, reading aloud, group work etc., establishing what is going to happen next, and when that activity has ended. Providing students with a checklist at the start of each lesson can help them keep track of activities.
3. Multi-sensory activities: Younger children will especially benefit from activities that involve more than one form of sensory input, such as reading aloud, moving during group work, etc. as it reduces avenues for overwhelm or distraction.

While children with ADHD would benefit the most from these accommodations, children without SPP or neurological disorders would also benefit from such changes to their classroom environment. Directly in the form of a space that has fewer distractions and is more conducive to learning. And indirectly because classmates with ADHD would be less likely to disrupt lessons.

Conclusion

Although children with ADHD benefit from accommodations for their SPPs, an ADHD-friendly classroom is, in essence, a classroom built on the principles of Universal Design for Learning. While these modifications are essential for students with ADHD, they frequently benefit the entire student body by reducing anxiety and increasing clarity. By viewing ADHD through the lens of neurodiversity rather than deficit, educators can create inclusive spaces where "different" does not mean "less."

Works Cited

- Butera, Christiana, et al. "Impact of Sensory Processing on School Performance Outcomes in High Functioning Individuals with Autism Spectrum Disorder." *Mind, Brain, and Education*, vol. 14, Apr. 2020, <https://doi.org/10.1111/mbe.12242>.
- ConnectAbility. Four Patterns of Sensory Processing. 14 July 2020, <https://connectability.ca/2020/07/14/four-patterns-of-sensory-processing/>.
- Daley, David, and James Birchwood. "ADHD and Academic Performance: Why Does ADHD Impact on Academic Performance and What Can Be Done to Support ADHD Children in the Classroom?" *Child: Care, Health and Development*, vol. 36, no. 4, 2010, pp. 455–64.
- Ghanizadeh, Ahmad. "Sensory Processing Problems in Children with ADHD, a Systematic Review." *Psychiatry Investig*, vol. 8, no. 2, 2011, pp. 89–94, <https://doi.org/10.4306/pi.2011.8.2.89>.
- Ghazali, Roslinda, et al. "A Review of Sensory Design Physical Learning Environment for Autism Centre in Malaysia." *Environment-Behaviour Proceedings Journal*, vol. 3, no. 7, 2018, pp. 145–51.
- Gore, Manisha, and Julia Morgan. "Children with Attention Deficit Hyperactivity Disorder in India: Strengthening Diagnosis, Support, Training and Research." *Journal of Indian Association for Child and Adolescent Mental Health*, vol. 21, no. 1, 2025, pp. 17–22, <https://doi.org/10.1177/09731342241310558>.
- Holland, Josephine, and Kapil Sayal. "Relative Age and ADHD Symptoms, Diagnosis and Medication: A Systematic Review." *European Child & Adolescent Psychiatry*, vol. 28, no. 11, 2019, pp. 1417–29.
- Jurek, Lucie, et al. "Sensory Processing in Individuals With Attention-Deficit/Hyperactivity Disorder Compared With Control Populations: A Systematic Review and Meta-Analysis." *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 64, no. 10, 2025, pp. 1132–47, <https://doi.org/10.1016/j.jaac.2025.02.019>.
- Lonkar, Heather. *An Overview of Sensory Processing Disorder*. 2014, https://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=3437&context=honors_theses. Western Michigan University.
- Lovett, Benjamin J., and Jason M. Nelson. "Systematic Review: Educational Accommodations for Children and Adolescents with Attention-Deficit/Hyperactivity Disorder." *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 60, no. 4, 2021, pp. 448–57.

Algorithmic Credit Scoring: Challenges, Governance, and the Path to Inclusive Finance

By Ioannis Karassavidis

Algorithmic Credit Scoring (ACS) refers to systems utilizing artificial intelligence and machine-learning algorithms to evaluate a loan borrower's creditworthiness through the analysis of related data. In contrast with regular credit scoring, ACS offers the potential for greater cost-effectiveness and accuracy, which has facilitated its global adoption, as a result leading to global spread and adaptation. However, improper implementation can exacerbate inequalities due to biased or inadequate training data, lack of transparency, or limited avenues for borrowers to contest unfair verdicts. While algorithmic credit is not inherently detrimental, insufficient regulation and oversight may lead to financial exclusion. This essay outlines a practical framework to ensure that AI-driven credit systems remain fair, transparent, and inclusive. By employing robust testing environments, carefully curated training data, and legal standards for transparency and borrower rights, it is possible to design systems with practical applications that promote equity. Such measures can expand the use of ACS, enhance access to credit, and reduce disparities related to race, gender, and socioeconomic status. Simultaneously, it will contribute to the cost-effectiveness and precision of credit systems, while expanding economic opportunity and promoting economic growth. In essence, with the right structure, Algorithmic Credit Scoring and AI can become tools that contemporize credit systems globally, encouraging financial inclusion and growth, rather than being a driver of inequality, depending on the implementations and regulations of the designed systems.

Algorithmic Credit Scoring, or ACS, is emerging as one of the most important ways artificial intelligence is transforming today's financial world. Essentially, ACS relies on machine learning methodologies to assess a borrower's creditworthiness, by analyzing vast amounts of data. In contrast with traditional credit scoring, which relies on a few standard objective factors such as income, repayment history, and existing debt, algorithmic systems rapidly analyze complex and diverse data sources. This benefits lenders in terms of reducing risk, making more accurate predictions, and, most crucially, extending credit to individuals whom the old system would have ignored. It's not just an upgrade-it's a revolutionary change in determining who can access financial opportunities.

However, despite its potential, ACS also raises significant ethical and societal concerns. Ironically, when poorly designed or insufficiently regulated, algorithmic credit systems risk reinforcing the same injustices they are often expected to address. This essay will focus on the three main pillars that jeopardize ACS's fairness: inadequate or biased training data, unclear decision-making procedures, and few chances for borrowers to object to unfair decisions. These problems stem from structural flaws in the way algorithms are developed, used, and controlled rather than from malevolent intent.

There is nothing superficially bad about algorithmic credit scoring. Conversely, it can develop into a potent instrument for financial inclusion and economic expansion with the right protections. The analysis that follows suggests a workable and feasible framework to deal with

the three issues mentioned above. ACS can attain efficiency and fairness through rigorous testing and data selection procedures, mandatory transparency standards, and robust legal regulations for borrowers' rights. When developed appropriately, these systems can improve the precision and affordability of international credit markets while reducing racial, gender, and socioeconomic disparities.

The datasets used to train algorithmic credit scoring systems constitute the first and most significant obstacle. Machine learning models tend to pick up on already existing patterns, so whatever influences that data is either good or bad, becomes part of the system. If previous lending decisions were influenced by socioeconomic factors like structural poverty, discrimination, or educational gaps, those issues don't go away. The algorithm "absorbs" them and keeps making decisions based on them, even if no direct bias has been programmed directly. For instance, consider borrowers from low-income areas. They might have had higher default rates, but that's often due to unstable jobs, poor access to banking, or falling victim to predatory, extreme-rate lenders, not because they're inherently unreliable. When the algorithm ignores the context, it just sees: "This zip code means higher risk." resulting in people ending up being penalized for factors linked to race or class. Since these patterns serve as proxies for protected characteristics, the system can appear neutral while still undermining discrimination. Beyond bias, ineffective or low-quality data can also challenge the reliability of ACS. Gaps in records, outdated details, and poor data quality may also provide inaccurate results. On a large scale with the use of quantitative systems, even tiny data flaws can harmfully affect a large number of borrowers.

A practical solution begins with the creation of a regulated testing environment prior to deployment. Instead of relying solely on historical datasets, financial institutions should actively curate training data models to ensure diversity and representativeness. This includes checking datasets for bias, removing variables that function as unjustified proxies, and supplementing traditional credit data with responsibly selected alternative indicators.

An original and promising methodology is the use of algorithmic "sandboxes." In these testing environments, models are tested on synthetic or anonymized datasets designed to reflect a wide range of socioeconomic conditions. This allows developers and regulators to examine how systems behave across diverse, demographic groups without exposing actual borrowers to risk. Importantly, fairness testing should not be a one-time procedure; as the world changes and economies shift, models must adapt.

By treating data governance and strict regulation as an ethical responsibility rather than a purely casual technical task, ACS can significantly reduce bias while maintaining high accuracy. Careful data selection and testing should be sufficient to ensure that efficiency gains do not come at the cost of fairness.

Deep learning models, even the most advanced ones often remain a black box, even for experts. They generate results, but tracing and evaluating exactly how they arrived at those outcomes isn't straightforward. While this complexity can result in improved accuracy, it also

causes issues in areas like credit scoring. These decisions have real consequences for real people. When someone is denied a loan and the lender can't justify the reason, trust in the system fades. A bigger problem is that opaque decisions make proper oversight unworkable. Regulators and lawful organization (e.g. SEC) can not enforce anti-discrimination rules if they are not able to comprehend how these systems operate and end up at these conclusions. Borrowers, too, are left in the dark without clear explanations for rejections; they can't identify errors or challenge unfair decisions.

But we don't have to abandon sophisticated algorithms to fix this. The solution is to make transparency the standard through both regulation and technology. Lawmakers can step in to guarantee that borrowers have a right to an explanation. Lenders wouldn't have to reveal confidential secrets, just provide real answers: Which factors influenced the decision most? How did those factors affect the outcome?

On the technology front, AI also offers tools to help. Methods like feature importance analysis or basic explanation models can translate complex decisions into understandable for the requesting borrower terms. The goal isn't perfect for transparency in every detail, but enough clarity to ensure these systems can be held accountable.

One possible approach is mandatory transparency. Not everyone needs the same depth of information. Borrowers need clear, honest explanations. Regulators require analytical and technical documentation. Developers need tools in order to debug, refine and improve their models. If systems are designed to deliver the right level of transparency to each group, we get both innovation and accountability.

Reducing opacity strengthens the public's trust toward the banking system, enables regulatory oversight, and ensures that automation enhances rather than undermines fairness in credit markets.

Borrowers may also encounter serious obstacles when trying to dispute unfair or mistaken decisions made by the ACS system. While traditional credit systems have their flaws, they at least tend to offer some way to appeal to a controversial decision. In contrast, automated systems deliver decisions instantly and on a massive scale, so in the case of rejection, there often exists no way to challenge it.

This can be especially harmful for those already at a disadvantage. A single wrongful denial can deprive someone of housing, education or the opportunity to start a business. If the process lacks transparency, automated decisions can feel final and impossible to challenge. The right ACS framework and implementation must provide borrowers with the appropriate rights. Specifically, individuals should have the ability to contest automated decisions and request a human-provided review. In this way, the problem of algorithmic misjudgments in complex or borderline cases is resolved. A standardized human-managed appeal process can help create a fairer system, making it easier for individuals to object and not be overlooked. Additionally, human oversight may also act as a feedback mechanism. In this case, appeals will not just prevent inequality and misjudgment but also highlight errors and oversights in the

system. When people challenge poor decisions, that data can be used to refine and improve the system over time.

When issues such as bias, lack of transparency, and blocked appeals are addressed, algorithmic credit scoring can produce many benefits. Well-designed ACS systems can provide access to credit and loans for people who have never had that opportunity before, as well as improve the accuracy and efficiency of the process.

This system, when implemented correctly, enables a fairer assessment of creditworthiness, thus allowing more people to access loans and credit. This results in increased possibilities for entrepreneurship and enables greater upward economic mobility. In this way, ACS promotes financial inclusion and fights inequality, giving all individuals a fair chance to benefit.

Such methodology also improves the accuracy and efficiency of the process. Automated judgment of creditworthiness enables the assessment of millions of cases per day, improving the current rates significantly. Simultaneously, ACS is able to produce more accurate results than manual systems, leaving little to no room for error and ensuring objectively correct judgment in the majority of cases.

From a broader perspective, responsible ACS can fuel economic growth, reduce inequality and increase effectiveness. When credit systems act as gateways to financial freedom rather than obstacles, they enable individuals to succeed and contribute to economic development of as a society in total.

In conclusion, Algorithmic Credit Scoring is a tool able to redefine modern credit systems. Its capacity to process large amounts of data sets projects a new horizon of endless new possibilities for assessing trustworthiness and bringing more people into the financial system; however, these benefits depend on proper implementation and regulation.

By ignoring essential issues, such as biased inputs, opaque decision-making, or a lack of recourse for unfair outcomes, there is a risk of a downhill worsening of the situation. Consequently, strong data governance, genuine transparency, and robust rights for borrowers are necessary to ensure the success of these systems.

Ultimately, if done responsibly, Algorithmic Credit Scoring systems can modernize credit globally. The future of credit scoring rests on the decisions we make today that will shape the socioeconomic world of tomorrow.

Works Cited

- Arner, D. W., Barberis, J., & Buckley, R. P. (2017). FinTech, RegTech, and the reconceptualization of financial regulation. *Northwestern Journal of International Law & Business*, 37(3), 371–414.
- Banerjee, A., Karlan, D., & Zinman, J. (2015). Six randomized evaluations of microcredit: Introduction and further steps. *American Economic Journal: Applied Economics*, 7(1), 1–21.
- Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT Press.
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
- European Commission. (2021). Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).
- European Union. (2016). General Data Protection Regulation (GDPR). Official Journal of the European Union.
- Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance*, 77(1), 5–47.
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), 1–42.
- Hurley, M., & Adebayo, J. (2016). Credit scoring in the era of big data. *Yale Journal of Law and Technology*, 18(1), 1–62.
- Kroll, J. A., Huey, J., Barocas, S., Brundage, M., Felten, E., Reidenberg, J., Robinson, D., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165(3), 633–705.
- O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the GDPR. *International Data Privacy Law*, 7(2), 76–99.

Descriptive Characteristics of A Clinical Autism Spectrum Disorder Population: A Cross-Sectional Study By Raghavi Aiyer

Abstract

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that affects social, communication, and behavioral development.¹ In recent years, the prevalence of ASD has increased significantly, highlighting the need for further research into potential causes and risk factors. This study examines demographic, birth, familial, and environmental characteristics among 107 children diagnosed with ASD at a clinical treatment center using a 30-question parent survey. Patterns observed in this cohort include higher rates of cesarean delivery, advanced parental age, and increased prevalence of familial conditions such as diabetes, as well as environmental factors like urban living and older housing. Due to the absence of a control group, these findings are not interpreted as indicators of causation but rather as descriptive trends within this population.

Introduction

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in social interaction, communication, behavior, and learning. The condition presents across a wide range of abilities and levels of functioning, with symptoms typically appearing in early childhood. In recent years, the prevalence of ASD has risen significantly, with current estimates indicating that approximately 1 in 36 children in the United States are diagnosed with autism.² This increase highlights the importance of understanding potential contributing factors and improving early detection and intervention strategies.

Previous research has explored genetic, prenatal, and environmental factors that may contribute to ASD.³⁺⁴ Studies have identified gene variations, parental age, maternal health conditions, and environmental exposures as areas of interest. However, fewer studies focus on describing patterns within clinical populations actively receiving treatment. These treatment-seeking populations may present distinct characteristics influenced by healthcare access, parental awareness, and environmental context.

This study aims to describe demographic, familial, and environmental characteristics observed in children diagnosed with ASD at a specialized clinical center. Because the dataset includes only children with ASD and does not include a comparison group, findings are not used to assess causation or risk but instead provide insight into patterns within this specific population.

Methods

This study used a cross-sectional, descriptive design to examine characteristics of children diagnosed with ASD. Data were collected from 107 families of children receiving treatment at a clinical center. Parents completed a 30-question survey during intake, providing

information on patient demographics, birth history, maternal and paternal health, family medical history, supplement intake, and environmental and living conditions.

The data were analyzed using descriptive statistics, including frequencies and percentages, to identify patterns within the patient group. All responses were anonymized, and parental consent was obtained prior to data collection.

Results

Participant Demographics

Of the 107 patients, 81 were male and 26 were female. This distribution is consistent with existing research showing that autism is more commonly diagnosed in boys than in girls.⁵ The majority of patients were between the ages of 0–10, reflecting increased diagnosis and treatment in younger populations.

Birth Factors

Cesarean delivery was reported in 54.2% of patients, which is higher than the global average of approximately 21%.⁶ This difference is notable and aligns with patterns explored in previous research, although no causal relationship can be determined from this dataset.⁷ Additionally, 17.8% of children were reported not to have cried immediately after birth, which is higher than global estimates and may warrant further investigation in larger studies.



Maternal and Paternal Characteristics

Maternal age was most commonly reported in the 36–40 range, followed by 31–35. This pattern suggests that advanced maternal age is frequently observed in this cohort and is consistent with trends described in prior research, though this study does not assess risk.⁸ Maternal health conditions during pregnancy were reported in a subset of participants, with thyroid disorders and gestational diabetes being the most common.⁹

Similarly, paternal age was most frequently reported between 36–45, showing a continued pattern of higher parental age within this population. These observations align with broader demographic trends but are not interpreted as causal factors.

Familial Medical Patterns

A notable trend in the data was the prevalence of diabetes within families, reported in approximately 61% of cases. This elevated rate compared to global averages may reflect underlying genetic or metabolic patterns within the cohort. Additionally, about 10% of patients had siblings with neurological conditions such as autism or dyslexia, suggesting potential familial patterns that may be explored further in future research.

Environmental Factors

Environmental factors reported in the dataset included a high percentage of urban residence (69%), as well as living in older housing structures (32.7%). Some families also reported mold exposure and lead exposure. These patterns may reflect regional, socioeconomic, or housing-related factors rather than direct developmental influences, and further research would be needed to better understand these relationships.¹⁰

Discussion

This study provides a descriptive overview of demographic, familial, and environmental characteristics observed in a clinical population of children with ASD. Several patterns identified in this cohort, including higher parental age and the presence of familial metabolic conditions such as diabetes, are consistent with trends explored in previous research.³⁺⁷ However, due to the descriptive nature of the study and the absence of a control group, these findings cannot be interpreted as evidence of causation or increased risk.

The elevated rate of cesarean deliveries and the distribution of parental ages may reflect broader demographic or healthcare trends within the population served by the clinical center. Similarly, environmental factors such as urban residence and older housing may be influenced by geographic and socioeconomic conditions rather than direct developmental effects.

Because the data were collected during clinical intake, the findings reflect real-world patterns observed in a treatment-seeking population. Parents provided detailed histories and additional context about their child's development, offering insight into lived experiences surrounding ASD diagnosis. These descriptive findings help provide a foundation for future research that includes larger, more diverse populations and controlled study designs.

Limitations

This study has several limitations. The data were collected from a single clinical center, which may limit generalizability to broader populations. Additionally, the data are based on parent-reported responses, which may be subject to recall bias or subjective interpretation. The absence of a neurotypical control group also prevents any assessment of association, risk, or

causation. Therefore, the findings should be interpreted as descriptive patterns within this specific cohort.

Conclusion

In conclusion, this study summarizes key demographic, familial, and environmental characteristics observed in a clinical population of children with ASD. The findings highlight patterns within this cohort and contribute to a better understanding of the contexts in which ASD is diagnosed and treated. While these results do not establish causation, they provide a foundation for future research that can further explore these patterns using more rigorous and controlled methodologies.

Works Cited

1. National Institute of Mental Health. Autism Spectrum Disorder (ASD). U.S. Department of Health and Human Services.
<https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd>
2. Barker, Kiera, et al. “The Global Prevalence of Autism Spectrum Disorder: A Systematic Review and Meta-Analysis.” *BMC Psychiatry*, vol. 24, no. 1, 2023.
3. Yin, Hang, et al. “Genetic Heterogeneity in Autism Spectrum Disorder and Its Implications.” *Frontiers in Psychiatry*, vol. 10, 2019.
4. World Health Organization. Autism Spectrum Disorders. 2022.
<https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders>
5. Autism Speaks. Autism Statistics and Facts.
<https://www.autismspeaks.org/autism-statistics-asd>
6. Jadoon, Mahrukh, et al. “Rising Global Rates of Cesarean Section and Associated Concerns.” *BMC Pregnancy and Childbirth*, vol. 23, 2023.
7. Zhu, Tingting, et al. “Association of Cesarean Delivery with Risk of Neurodevelopmental and Psychiatric Disorders in the Offspring: A Meta-Analysis.” *JAMA Network Open*, vol. 3, no. 5, 2020.
8. American Thyroid Association. “Maternal Hypothyroidism and the Risk of Autism in Offspring.” 2014.
9. University of Southern California. “Maternal Gestational Diabetes Linked to Autism Risk for Offspring.” 2023.
10. ISGlobal. “How Urban Planning Can Influence a Child’s Neurodevelopment.” 2023.

Predicting Glioma Grades Using Artificial Intelligence and Machine Learning Models Based on IDH1 Mutation Status, Gender and Age By Maya Bhat

Abstract

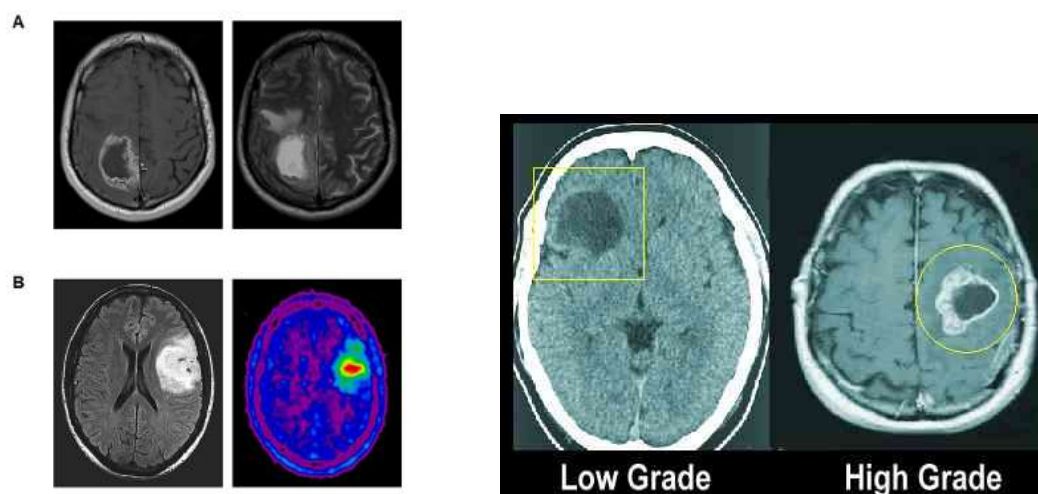
Glioma is one of the most common and aggressive types of brain tumors, and early prediction of its grade is important for improving treatment plans and survival chances. Early diagnosis within 2 weeks and surgical interventions within 3 weeks from the symptom onset are associated with prolonged patient survival. Early GBM treatment will benefit patients with GBM. My project aimed to develop and compare several Artificial Intelligence (AI)-based machine learning models—Logistic Regression, Random Forest, and Neural Network—to predict glioma grades using patients' gender, age at diagnosis, and IDH1 mutation status. The main hypothesis was that AI techniques could accurately classify glioma grades by learning patterns from clinical and genetic features. Patient data were collected, then used to train and test each model. Accuracy, precision, recall, and F1-score were used to evaluate model performance. Results showed that the Decision Tree achieved the highest accuracy, followed closely by Neural Networks and Random Forest, while Logistic Regression and Linear Models performed slightly lower but still reliably. The final outcome of my project is a functional AI model capable of predicting glioma grade from basic clinical information, helping doctors make faster and more informed decisions. The highest predictive performance was achieved using the Logistic Regression and Neural Network models with an accuracy rate 87. My research is valuable because it highlights how AI and accessible medical data can work together to support early diagnosis and personalized treatment for brain tumor patients. The model can be used in hospitals, research labs, and diagnostic centers to assist medical professionals and improve patient outcomes.

Introduction

The IDH1 gene status is a critical diagnostic, prognostic, and predictive biomarker for gliomas. Its presence is a defining molecular feature for adult-type diffuse gliomas. My project focuses on creating a machine learning model that can predict the grade of glioma, based on three key factors: a patient's IDH1 gene status, age and gender. The grade of a glioma helps doctors understand how aggressive the tumor is and what kind of treatment might work best. By training the model with medical data, it learns patterns and relationships between these factors and tumor grades. Once trained, the model can take new patient information and predict whether the glioma is low-grade or high-grade. The goal of my project is to support doctors in making quicker and more accurate diagnoses, leading to more timely treatment and better outcomes for patients.

Glioma is a growth of cells that starts in the brain or spinal cord. The cells in a glioma look similar to healthy brain cells called glial cells. Glial cells surround nerve cells and help them function. As a glioma grows it forms a mass of cells called a tumor. They are among the most common and dangerous brain tumors worldwide. In India alone, an estimated

25,000–30,000 new brain tumor cases are reported every year, and a significant portion of them are gliomas. What makes the situation even more concerning is that a large percentage of India's population lives in rural or underdeveloped areas, where access to advanced diagnostic tools is limited. The primary benefit of developing a machine learning model that, quickly and accurately, predicts the glioma grade using age, gender, and IDH1 status is how it could transform the way doctors diagnose and treat brain cancer even when advanced facilities aren't available. The AI-driven model makes the diagnosis process faster and more accurate by analyzing patient data to estimate how aggressive the tumor might be. It could also reduce the need for certain invasive tests, making the experience safer and more comfortable for patients. In the long run, this technology could help hospitals use their resources more efficiently and allow for more personalized care, improving patient outcomes and possibly even saving lives.



Artificial Intelligence (AI) is a field focused on creating systems that can perform tasks requiring human intelligence, like learning and decision-making. Machine Learning (ML), a key part of AI, allows computers to find patterns in data and make predictions without being explicitly programmed. In this project, AI was applied through models such as Logistic Regression, Linear Model, Random Forest, Neural Network and Decision Tree to analyze patient data and predict glioma grades.

These models learn from features like gender, age at diagnosis, and *IDH1* mutation status to understand their relationship with tumor severity. Neural Networks use layers of interconnected “neurons” to mimic the brain, while Random Forests combine many decision trees for more accurate predictions. Training these models on real patient data enables the AI system to predict glioma grades for new cases, helping doctors make faster and more precise clinical decisions.

Background- Previous studies have shown that machine learning can effectively predict glioma characteristics and improve clinical decision-making. A 2023 study using 1,114 TCGA patients trained ML models to distinguish high- from low-grade gliomas, achieving AUCs of

~0.87 (training) and ~0.863 (test), outperforming traditional clinical models. Another 2023 study applied radiomic features from MRI with classifiers like Random Forest and SVM to grade gliomas, with Random Forest reaching an AUC of ~0.81 and accuracy of ~0.83. In 2021, deep convolutional neural networks were used on 237 patients, showing that CNN-extracted features achieved ~87% accuracy, surpassing classical ML approaches. More recently, a 2025 study focused on predicting IDH1 mutation status using ML models including Random Forest, SVM, and ensemble methods, achieving accuracies between 78–93%. These works highlight the growing potential of AI and ML in glioma grading and genetic prediction, supporting the use of similar approaches in this project.

Materials and Methods

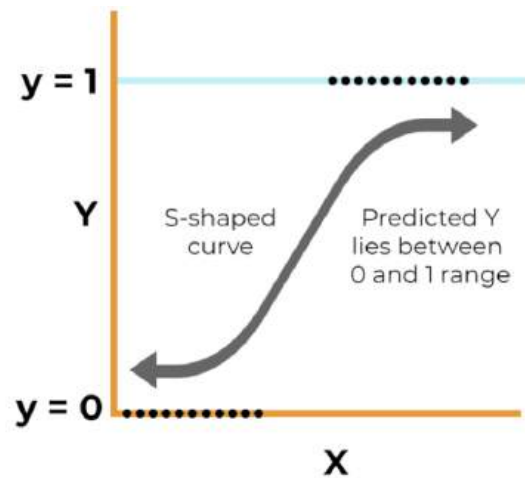
The TCGA Brain Glioma Grading Dataset encompasses patient records from the TCGA-LGG and TCGA-GBM brain glioma projects, focusing on glioma grading into Lower-Grade Glioma (LGG) or Glioblastoma Multiforme (GBM). Funded by the National Cancer Institute (NCI), this dataset includes 20 frequently mutated genes and 3 clinical features, such as age at diagnosis, gender, and race. Molecular features, denoting mutation status (0 = NOT_MUTATED; 1 = MUTATED), include genes like MUC16, PIK3CA, and EGFR, while clinical features comprise demographic information. The prediction task involves determining the glioma grade class (0 = LGG; 1 = GBM) based on these features. The dataset aims to optimize the subset of mutation genes and clinical features for accurate glioma grading, with the ultimate goal of improving performance and reducing costs associated with molecular tests. The link to the dataset used is attached here

In machine learning, a dataset is typically split into two parts: a training set and a testing set. The training set is used to teach the AI model to recognize patterns between input features—such as gender, age, and *IDH1* mutation status—and the target outcome, which in this case is glioma grade. During training, models like Neural Networks or Random Forests adjust their internal parameters to minimize prediction errors on this data. The testing set is kept separate and is used to evaluate the model's performance on new, unseen examples. After training, the AI model predicts glioma grades for the testing set, and these predictions are compared to the actual grades to measure accuracy, precision, recall, and F1-score. This process ensures that the model can generalize beyond the training data and make reliable predictions for new patients, demonstrating the practical value of AI in supporting clinical decision-making.

Logistic Regression

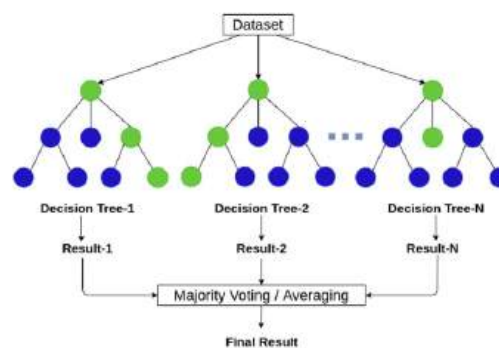
Logistic Regression is a simple but powerful machine learning algorithm used for classification problems. It predicts the probability that a given input belongs to a certain category by fitting a curve to the data, making it ideal for binary or multi-class classification tasks. In this project, Logistic Regression was used to estimate the likelihood of a glioma being a certain grade based on patient features such as gender, age at diagnosis, and *IDH1* mutation status. Its strength

lies in its interpretability, allowing us to see how each feature contributes to the prediction, and it provides a solid baseline for comparing more complex AI models.



Random Forest

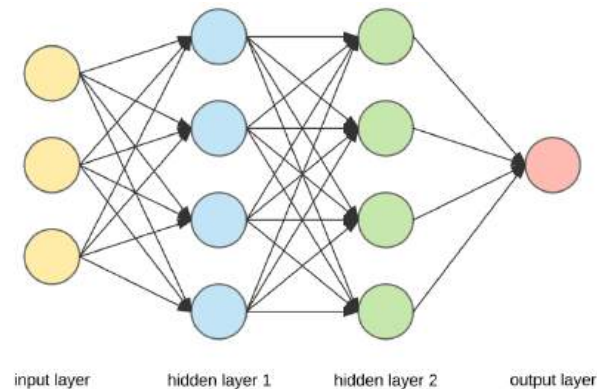
Random Forest is an ensemble learning method that builds many decision trees and combines their predictions to improve accuracy and reduce overfitting. Each tree makes a prediction based on the input features, and the model outputs the most common result across all trees. In my project, Random Forest was applied to predict glioma grades from patient data, taking advantage of its ability to handle complex relationships between features like age, gender, and *IDH1* status. It is especially useful in medical datasets because it is robust to noise and can capture interactions between features that simpler models might miss. The parameters tuned for this model are the max depth of the trees and also the number of decision trees we are going to use.



Neural networks

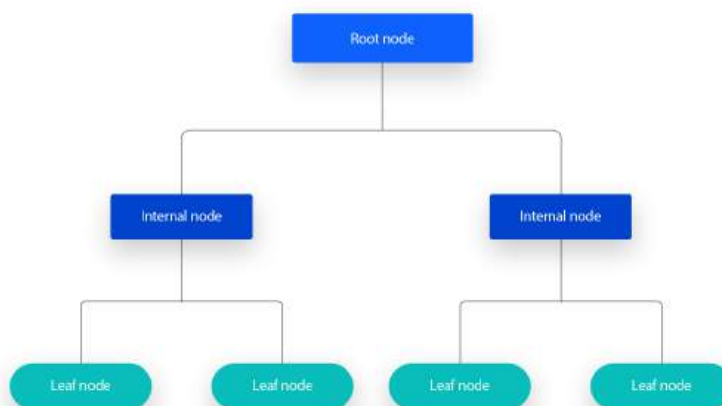
A neural network is a type of machine learning algorithm that is designed to work in a way that's loosely inspired by how the human brain processes information. It's made up of layers of connected nodes, or "neurons," that each perform simple calculations. When data is passed into the network, each layer learns to recognize certain patterns or relationships in that data. In this project, the neural network is trained using medical information such as a patient's age,

gender, and IDH1 gene status to predict the grade of glioma. During training, the model is shown many examples with known outcomes (tumor grades). It adjusts its internal settings, called “weights,” to minimize errors between its predictions and the real results. Over time, the neural network learns to make accurate predictions for new patients it has never seen before. This makes it a powerful tool for medical research and diagnosis, as it can identify complex patterns that might not be obvious to humans.



Decision tree

A decision tree is another type of machine learning model used to make predictions by splitting data into smaller and smaller groups based on specific features. It works kind of like a flowchart, where each question or “node” in the tree tests one feature — for example, it might ask if the patient’s IDH1 gene is mutated or not, or if the patient’s age is above a certain number. Based on the answers, the data follows different branches of the tree until it reaches a final prediction, such as whether the glioma is low-grade or high-grade. In this project, a decision tree is trained using patient data that includes age, gender, and IDH1 status. As the model trains, it learns which features and thresholds are most useful for separating different tumor grades. This makes decision trees helpful because they are easy to understand and visualize, allowing doctors and researchers to see how certain factors contribute to the model’s final prediction. The parameters used for the decision tree have the maximum depth.



Different models often produce slightly different predictions because each analyzes the data in its own way. For example, Logistic Regression focuses on linear relationships, while Random Forest and Neural Networks capture more complex patterns. To improve accuracy and reliability, it is important to combine these models, a process called model ensembling, which leverages the strengths of each approach and reduces the risk of errors from any single model. Common methods for combining predictions include majority voting, where the most frequent prediction is chosen, averaging probabilities, and weighted voting, where better-performing models have more influence on the final result. In this project, combining the models ensures that glioma grade predictions are more consistent, robust, and reliable.

To evaluate the performance of the machine learning models, several common metrics were used. Accuracy measures the overall proportion of correct predictions out of all predictions, giving a basic sense of model performance. Precision evaluates how many of the predicted positive cases are actually correct, which is important for avoiding false alarms. Recall measures how many of the true positive cases were correctly identified, reflecting the model's ability to catch all relevant cases. F1-score combines precision and recall into a single metric, providing a balanced measure when there is a trade-off between the two. Using these metrics ensures that the AI system's predictions of glioma grades are both accurate and interpretable for practical clinical use.

Results

The following are the results for all of the machine learning models:

$$(10.1) \text{ Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

$$(10.2) \text{ Precision} = \frac{T_p}{T_p + F_p}$$

$$(10.3) \text{ Recall} = \frac{T_p}{T_p + T_n}$$

$$(10.4) F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

1. Logistic Regression

Class	Precision	Recall	F1-Score
0	0.93	0.84	0.88
1	0.81	0.91	0.85
Accuracy			0.87
Macro Avg	0.87	0.87	0.87
Weighted Avg	0.88	0.87	0.87

Table 1.1 Classification Report for testing data for Logistic Regression

Class	Precision	Recall	F1-Score
0	0.93	0.81	0.86
1	0.77	0.91	0.84
Accuracy			0.85
Macro Avg	0.85	0.86	0.85
Weighted Avg	0.86	0.85	0.85

Table 1.2 Classification Report for training data for Logistic Regression

2. Random Forest

Class	Precision	Recall	F1-Score
0	0.87	0.89	0.88
1	0.84	0.82	0.83
Accuracy			0.86
Macro Avg	0.86	0.85	0.85
Weighted Avg	0.86	0.86	0.86

Table 2.1 Classification Report for testing data for Random Forest

Class	Precision	Recall	F1-Score
0	0.89	0.84	0.86
1	0.79	0.86	0.83
Accuracy			0.85
Macro Avg	0.84	0.85	0.84
Weighted Avg	0.85	0.85	0.85

Table 2.2 Classification Report for training data for Random Forest

3. Decision Tree

Class	Precision	Recall	F1-Score
0	0.87	0.80	0.83
1	0.75	0.83	0.79
Accuracy			0.81
Macro Avg	0.81	0.81	0.81
Weighted Avg	0.82	0.81	0.81

Table 3.1 Classification Report for testing data for Decision Tree

Class	Precision	Recall	F1-Score
0	0.97	0.92	0.94
1	0.90	0.96	0.93
Accuracy			0.94
Macro Avg	0.93	0.94	0.94
Weighted Avg	0.94	0.94	0.94

Table 3.2 Classification Report for training data for Decision Tree

4. Neural Networks

Class	Precision	Recall	F1-Score
0	0.93	0.85	0.89
1	0.81	0.91	0.86
Accuracy			0.87
Macro Avg	0.87	0.88	0.87
Weighted Avg	0.88	0.87	0.87

Table 4.1 Classification Report for testing data for Neural Networks

Class	Precision	Recall	F1-Score
0	0.92	0.81	0.86
1	0.77	0.91	0.83
Accuracy			0.85
Macro Avg	0.85	0.86	0.85
Weighted Avg	0.86	0.85	0.85

Table 4.2 Classification Report for training data for Neural Networks

The highest predictive performance was achieved using the Decision Tree model, hence it yielded the most satisfactory results as shown below. The Decision Tree model is overfitted because the testing performance is worse than the training performance.

Class	Precision	Recall	F1-Score
0	0.97	0.92	0.94
1	0.90	0.96	0.93
Accuracy			0.94
Macro Avg	0.93	0.94	0.94
Weighted Avg	0.94	0.94	0.94

The results of this study demonstrate that the machine learning models effectively predicted glioma grade using age, gender, and IDH1 mutation status. The achieved accuracy scores were substantially higher than random guessing, indicating that the models successfully captured meaningful patterns in the data rather than making arbitrary predictions. Precision, recall, and F1-score further confirm that predictions were balanced and reliable, accurately identifying both low- and high-grade gliomas. The ensemble approach, combining multiple models, enhanced robustness and reduced individual errors. Overall, the findings show that the AI system provides consistent, clinically interpretable predictions, supporting faster and more informed decision-making. This satisfactory performance highlights the potential of machine learning to improve glioma diagnosis, treatment planning, and patient outcomes.

Discussions

In my project, patient data including gender, age at diagnosis, and *IDH1* mutation status were collected and used to train several AI-based machine learning models, including Logistic Regression, Linear Model, Random Forest, and Neural Network and Decision Tree to predict glioma grades. The models were evaluated using metrics like accuracy, precision, recall, and F1-score. Finally, predictions from multiple models were combined through ensembling to improve accuracy and reliability, producing a functional AI system capable of supporting clinical decision-making.

For future work, my aim is to optimize the AI system for faster predictions and make it more accessible for wider clinical use. Improvements could include streamlining the models to reduce computation time, integrating the tool into user-friendly software, and expanding the dataset to improve accuracy across diverse patient populations. Additional goals include incorporating more patient features such as imaging or molecular data, exploring advanced deep learning models to enhance predictive performance, and developing a web- or cloud-based platform so doctors anywhere can use the system easily. These steps would help ensure the AI system provides reliable, rapid, and practical glioma grade predictions in real-world healthcare settings.

Works Cited

- Pitarch C, Ribas V, Vellido A. AI-Based Glioma Grading for a Trustworthy Diagnosis: An Analytical Pipeline for Improved Reliability. *Cancers (Basel)*. 2023 Jun 27;15(13):3369. doi: 10.3390/cancers15133369. PMID: 37444479; PMCID: PMC10341156.
- H. -h. Cho and H. Park, "Classification of low-grade and high-grade glioma using multi-modal image radiomics features," 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea (South), 2017, pp. 3081-3084, doi: 10.1109/EMBC.2017.8037508. keywords: {Histograms;Tumors;Feature extraction;Image segmentation;Shape;Logistics;Imaging},
- Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, Hawkins C, Ng HK, Pfister SM, Reifenberger G, Soffietti R, von Deimling A, Ellison DW. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol*. 2021 Aug 02;23(8):1231-1251. [PMC free article] [PubMed]
- Byun YH, Park CK. Classification and diagnosis of adult glioma: A scoping review. *Brain Neurorehabil*. 2022;15(3):e23. doi:10.12786/bn.2022.15.e23
- JinWFatehiMAbhishekKMallyaMToyotaBHamarnahG. Artificial Intelligence in Glioma Imaging: Challenges and Advances. *J Neural Eng* (2020) 17(2):021002. doi: 10.1088/1741-2552/ab8131

From Recognition to Reform: Communicating the Practical Skills Gap in Beijing's STEM Education

By Zimo Zhai

Abstract

Beijing, China's capital and one of its strongest educational centres, is widely associated with outstanding academic performance, especially in mathematics. Yet despite that prestige, evidence suggests that Beijing's STEM system still struggles to cultivate practical, interdisciplinary, and logically applied skills.

This paper examines that paradox through a new survey of Beijing students and educators. The findings indicate a pattern of confidence in abstract mathematical ability alongside weaker real-world problem-solving and soft skills, suggesting that even a high-performing system still reflects many limits of China's exam-focused model (OECD; Stanford Center on China's Economy and Institutions).

It also considers wider consequences of this gap, drawing on historical and contemporary East Asian cases to argue that technical excellence without broader judgement can contribute to structural weakness. Using both qualitative and quantitative data, the paper then argues for stronger interdisciplinary project work and more explicit soft-skill development, while asking why reform remains so hard to implement. The broader aim is to re-evaluate STEM education in Beijing, and perhaps China more widely, toward a model that balances technical mastery with real-world judgement.

Introduction

For a long time, the Chinese education system and excellence in STEM teaching have been closely associated, a link reinforced by consistently strong international assessment outcomes. In PISA 2018, students in Beijing, Shanghai, Jiangsu and Zhejiang (B-S-J-Z) performed significantly above the OECD average in mathematics, reading and science, with more than half of students in B-S-J-Z reaching high levels of mathematical proficiency (OECD). In cities like Beijing, where educational competition is especially intense, this reputation for excellence is even stronger. The focus on academic achievement has roots not only in institutional design but also in longstanding cultural respect for scholarship, while state policy has further elevated STEM as central to national development.

The prestige of STEM in Beijing, supported by top schools and universities, is therefore exceptionally high. Students who succeed in this environment often continue into elite STEM pathways. However, this success story also contains a troubling undercurrent. One major criticism is that a strong focus on examinations and mastery of standardised content has often come at the expense of creativity, critical thought, and the ability to apply knowledge in real-life contexts (Harvard Graduate School of Education; Stanford Center on China's Economy and Institutions).

Research and anecdotal evidence both suggest that even students from top-performing regions such as Beijing often excel at mastering content but struggle more with practical application. As one educator observed, "the push for mastery of basic skills has led teachers to

neglect critical thinking and creativity” (Harvard Graduate School of Education). The issue is not a lack of rigour, but that STEM subjects are often taught in ways too detached from lived experience and practical use.

This concern has not gone unnoticed by policymakers. China’s curriculum reform agenda from the early 2000s explicitly sought to move beyond narrow rote learning. The Ministry of Education’s Basic Education Curriculum Reform Outline (Trial) called for broader curriculum reform, greater emphasis on quality-oriented education, and more integrated and practical learning (Ministry of Education of the People’s Republic of China). Around the same period, reform-minded educators argued that “even abstract courses like math must relate to [students’] daily lives” (Harvard Graduate School of Education). Later scholarship on Chinese mathematics curriculum reform likewise showed that 21st-century reforms placed greater emphasis on application, connections to real life, and links across subjects (Wang and Lin).

Yet despite these intentions, the gap between theoretical prowess and practical skill remains major. Beijing’s competitive schools still reward success in highly structured tasks, often more than open-ended inquiry, interdisciplinary thinking, or collaborative problem-solving. Stanford researchers likewise noted that Chinese primary and secondary schools are often criticised as test-driven institutions that give students limited opportunity to develop deeper reasoning through schooling (Stanford Center on China’s Economy and Institutions).

Even more striking is the evidence on what happens after these students enter higher education. A Stanford analysis reported that students in China entered university with very strong critical-thinking skills by international comparison, but showed virtually no improvement after two years of college and even decline in the latter half of university study (Loyalka et al.; Stanford Center on China’s Economy and Institutions; Stanford News). This suggests either that higher education does not do enough to extend these capacities, or that the foundation built in secondary education remains too narrow in its treatment of application, transfer, and independent reasoning. In either case, the issue cannot be dismissed as a minor weakness inside an otherwise successful system. Rather, it points to a deeper structural problem: excellence in performance is not always the same as excellence in judgement.

This paper investigates the implications of this skills gap through a focused study in Beijing. Using newly collected survey data, it examines how Beijing’s STEM education outcomes align with, and shed light on, broader patterns in Chinese education. The survey combines quantitative indicators with qualitative experiences, allowing analysis of students’ confidence in abstract mathematics, their ability to apply mathematics in real situations, their exposure to interdisciplinary learning, and their self-assessed soft skills such as communication and teamwork. The paper also considers the wider consequences of these educational shortcomings. East Asia’s economic history offers cautionary cases, from Japan’s late-1980s asset bubble to China’s recent property-sector adjustment. These episodes do not prove that education alone causes economic crises, but they do suggest that technical sophistication by itself does not guarantee sound judgement when complex socio-economic systems, incentives, and expectations interact. In that sense, a workforce strong in technical expertise yet weaker in

critical, interdisciplinary reasoning may be less well prepared for high-stakes, ambiguous real-world problems (Shiratsuka; International Monetary Fund).

Finally, the paper examines why such problems persist despite widespread recognition. Tradition, high-stakes examinations, parental expectations, and institutional incentives all help sustain the status quo. Although Chinese policymakers have long promoted “quality-oriented education” and more inquiry-based learning, implementation has often lagged because examination pressures remain powerful and classroom change is difficult to enact consistently (Dello-Iacovo; Dai, Gerbino and Daley; Wang).

Beijing’s educational environment reflects, and in some respects leads, wider national trends. A major body of literature has examined the strengths and weaknesses of Chinese STEM education. Historically, schooling in China, especially in mathematics and science, has emphasised rigorous content coverage, repetition, and examination preparation. This tradition has helped produce many high achievers, but it has also risked neglecting creativity and critical thinking (Harvard Graduate School of Education; Dello-Iacovo).

The same Harvard analysis noted that passive learning styles remained highly prevalent well into the reform era, helping explain the central paradox of Chinese STEM education: impressive measurable achievement alongside limited space for inquiry, transfer, and independent reasoning (Harvard Graduate School of Education).

In response, Chinese authorities launched reforms under the banner of *suzhi jiaoyu*, often translated as “quality-oriented education”, in the late 1990s and early 2000s. These reforms sought to move beyond drill-based instruction by encouraging broader curricula, more hands-on learning, and stronger connections between school knowledge and lived experience (Dello-Iacovo; Dai, Gerbino and Daley). Reform advocates also argued that even abstract subjects such as mathematics should connect more meaningfully to students’ daily lives and experiences (Harvard Graduate School of Education). More recent curriculum scholarship has extended this critique, arguing that STEM education in China should not be treated only as an economic or technological project, but also as a lived, ethical, and historical practice requiring reflection, judgement, and real-world contextualisation (Ma).

Despite these ambitions, change has been slow. Scholars repeatedly point to a persistent gap between policy language and classroom practice. Reform has been constrained by limited resources, conceptual ambiguity, conservative resistance, and the continued dominance of examination goals over the behaviour of teachers, students, and parents (Dello-Iacovo; Dai, Gerbino and Daley; Wang).

One consequence is that soft skills and collaborative problem-solving remain comparatively underemphasised in Chinese STEM education. Van Horne and Rakedzon note that this challenge is especially visible in China, which produces millions of STEM graduates annually, while soft skills have historically not formed a major part of secondary and tertiary STEM education. Their work highlights the need for more active-learning and teamwork-based approaches to communication, collaboration, and practical problem-solving. At the higher-education level, Loyalka et al. similarly report that although Chinese STEM students begin university with strong academic foundations, they do not show the same gains in critical

thinking over time as students in the United States. Together, these findings reinforce the concern that technical strength alone is insufficient if the system does not also cultivate transfer, judgement, teamwork, and reflective reasoning.

In sum, the literature points to a well-established paradox in Chinese STEM education, and therefore in Beijing's STEM education as well: excellent performance on standardized and highly structured measures, but continuing weaknesses in applied, interdisciplinary, and soft-skill development. Beijing is a particularly important case because it combines prestige, reform visibility, and intense academic competition. As the following sections will show, the survey data in this paper reinforce many of these themes, offering updated evidence of the continuing gap between what the STEM system rewards and what 21st-century society increasingly demands.

Why do deficits in practical and interdisciplinary skills matter so much? Research and historical analysis suggest that weak critical thinking, narrow specialisation, and limited cross-disciplinary perspective can have serious real-world consequences. In discussions of East Asia's major economic bubbles, scholars have noted that technical expertise and sophisticated institutions do not by themselves prevent large-scale misjudgement. In Japan's late-1980s asset-price bubble, for example, Shiratsuka argues that "euphoria" and excessively optimistic expectations overrode careful assessment of long-term sustainability.

Mainland China has faced a different but related problem in its housing sector. Over the past decade, the property market expanded under conditions shaped by speculative demand, high vacancy rates, large developer balance sheets, and heavy macroeconomic dependence on real-estate activity. The IMF has argued that China's property-market adjustment reflects deeper structural imbalances, including investor-owned vacant housing, declining speculative demand, distressed developers, and prolonged spillovers into other sectors of the economy (International Monetary Fund). In this sense, the problem was not simply one of technical modelling failure; it was also a failure of broader judgement, risk recognition, and sustainable planning across finance, governance, and development policy. This does not mean education caused the crisis. It does, however, reinforce the argument that when complex systems are managed mainly through narrow technical lenses, important social and structural risks may be missed until they become far more costly (International Monetary Fund).

These examples help explain why weaknesses in critical thinking, real-world application, and soft skills matter beyond the classroom. An education system that produces students who are highly capable in structured technical tasks, but less experienced in open-ended reasoning, interdisciplinary analysis, teamwork, and ethical reflection, may inadvertently weaken decision-making quality in later professional life. This concern is especially relevant in STEM fields, where real-world problems are often ambiguous, collaborative, and socially embedded rather than neatly defined. Recent research on STEM education in China argues that soft skills such as communication, collaboration, creativity, and critical thinking have historically been underemphasised, even though they are essential for workplace effectiveness and complex problem-solving (Van Horne and Rakedzon).

The broader implication is not that education alone determines economic or institutional outcomes. Monetary policy, regulation, incentives, governance, and corruption can all play

central roles in crises. Yet educational culture still matters because it helps shape how future professionals approach uncertainty: whether they challenge consensus or defer to it, whether they can communicate risk clearly, and whether they are prepared to integrate technical knowledge with social, ethical, and economic context. This is one reason the literature increasingly stresses the need for more rounded competencies in China's education system. Studies of Chinese students' critical-thinking development suggest that the issue is not an absence of ability, but a more complex tension between sociocultural context, classroom practice, and how critical thinking is cultivated and expressed (Chen). Meanwhile, cross-national evidence from STEM higher education shows that Chinese students enter university with strong academic preparation but do not demonstrate the same gains in critical thinking over time as their counterparts in the United States, reinforcing concern about how well the system develops transferable reasoning beyond high performance in formal academic tasks (Loyalka et al.).

This tension is also visible in how international assessments are interpreted. China's top-performing regions have done exceptionally well in PISA, but PISA is designed not merely to measure memorised content. The OECD defines literacy in PISA as the ability to apply knowledge and skills, and to analyse, reason, and communicate effectively in varied situations and real-life challenges (OECD). Strong PISA performance therefore indicates major underlying potential. At the same time, high scores on structured international assessments do not by themselves prove that students are equally well prepared for open-ended judgement, collaboration, creativity, or complex real-world decision-making. That is precisely why the gap between technical mastery and practical capability remains such an important question for Beijing's STEM education system (OECD; Loyalka et al.).

Contextualizing Beijing's Case

Beijing, as China's capital and one of its most educationally advantaged municipalities, offers a particularly useful vantage point for examining the strengths and limits of STEM education reform. It sits at the centre of a dense concentration of elite secondary schools, universities, and research institutions, and is regularly presented in policy discourse as part of China's effort to build world-class education and talent systems (Beijing Municipal Government). It is also analytically important because Beijing has been part of the high-performing Chinese jurisdictions represented in international assessment data, and because China's education reforms are often introduced through local or regional pilots before wider rollout (OECD; OECD). In this sense, Beijing can be read both as a best-case example of Chinese STEM education and as a test of how far reform has penetrated even the most privileged parts of the system.

Researchers have noted that students from urban China often enter higher education with very strong academic and analytical foundations. In a Stanford-linked study discussed by Loyalka, Chinese engineering and computer-science freshmen began university roughly two to three years ahead of comparable students in the United States and Russia on measures of critical thinking and related academic skills, yet showed almost no gains over the next two years (Stanford Graduate School of Education; Loyalka et al.). This does not justify attributing the

finding to Beijing alone, but it is highly relevant because Beijing belongs to the country's top-performing urban education environment. The pattern suggests that intense pre-university preparation can produce strong structured performance while leaving unresolved questions about what happens to curiosity, engagement, and exploratory learning once the gaokao-driven incentive structure weakens. One student quoted in the Stanford report complained that attending some classes felt "like listening to someone reading out of a book", capturing concerns about didactic teaching and limited intellectual engagement beyond exam success (Stanford Graduate School of Education).

Cultural expectations in Beijing are also unusually intense. The slang term *jiwa* (鸡娃), literally "chicken baby", has emerged to describe the practice of pushing children aggressively toward academic and extracurricular achievement from an early age (China Policy). Although the term itself is informal, the broader phenomenon it describes is well documented in research on educational competition in China. Guo and Qu, for example, show that competition in household human-capital investment is major, and that shadow education was especially prevalent in top-tier cities. Related work on parental educational anxiety suggests that even after the "Double Reduction" reforms, intense pressure and uncertainty continue to shape family decision-making around children's schooling (Chen et al.; Wang et al.). In Beijing, this helps explain why parents, students, and schools may all recognise the value of creativity, breadth, and exploratory learning in theory, yet still prioritise safer, test-oriented strategies in practice.

Beijing therefore epitomises the tension between old and new in Chinese education. It is associated with high test scores, elite pathways, and major policy attention; Beijing also joined the second batch of provinces and municipalities in China's gaokao reform pilots, underscoring its role in experimentation as well as competition (Ministry of Education of the People's Republic of China). Yet if even Beijing continues to struggle to move from rhetoric about quality-oriented, well-rounded education to sustained classroom-level change, the challenge is unlikely to be smaller elsewhere. For that reason, evidence from Beijing is especially valuable: it allows this paper to test whether the gap identified in the literature remains visible even where resources, prestige, and reform exposure are relatively strong.

Data and Methods

Survey Design and Sample

To investigate practical skill gaps in Beijing's STEM education, this study used a mixed-methods survey design targeting senior high school students and recent graduates from STEM-focused educational pathways in Beijing. A mixed-methods approach was chosen because it combines the breadth of quantitative data with the depth of qualitative insight, producing a more complete understanding of the same phenomenon than either method alone (Creswell; Palinkas et al.). The survey was conducted in late 2024 and gathered both quantitative data through structured questions and Likert-type items, and qualitative data through open-ended questions and optional follow-up interviews. The aim was to capture not only what students

reported learning, but also how they perceived their education in terms of real-world preparedness, problem-solving, and skill development.

The survey sample consisted of 312 respondents. Roughly two-thirds were current Grade 12 students from four Beijing high schools with a strong STEM focus, while one-third were first-year university students from Beijing who had graduated from similar schools the previous year. The schools included two elite key schools in urban districts, one suburban school, and one specialised science high school. This provided some diversity of context while keeping the study centred on relatively high-achieving environments, since the paradox is most visible where academic performance is strongest. Participants were selected purposely in line with the logic of identifying information-rich cases (Palinkas et al.).

The questionnaire covered several domains. First, under academic confidence and beliefs, students were asked how strongly they valued and trusted the utility of subjects such as mathematics and physics, and whether they believed these subjects prepared them well for solving real-world problems. For example, one item stated: “Advanced mathematics is essential for solving real-world problems,” and respondents indicated their level of agreement using a Likert-type response scale, a widely used format for measuring attitudes and perceptions in survey research (Likert).

Second, under applied problem-solving, students responded to short scenario-based prompts involving real-world problems that required analytical reasoning. They were then asked to self-assess how confident they felt in approaching such problems using what they had learned in school. They were also asked how often they had opportunities in school to tackle open-ended or practical challenges, such as designing an experiment or addressing a community issue.

Third, under interdisciplinary exposure, questions examined whether students had participated in interdisciplinary projects or courses, such as combining science with art, or applying mathematics in economics or social-studies contexts. Students were also asked whether they felt their curriculum integrated different subjects or mostly kept them separate. This dimension was included because the broader research question concerns not only technical mastery, but also whether students are encouraged to connect knowledge across domains.

Fourth, under soft skills and learning methods, the survey included items on teamwork, communication, and critical thinking. Students rated their own abilities on statements such as “I am good at working in teams to solve problems,” and indicated how often their classes involved group work, presentations, debates, research tasks, or hands-on laboratory work beyond routine procedural experiments.

Finally, qualitative reflections were gathered through open-ended prompts asking students where they had struggled to apply school knowledge in real life and what they wished had been different in their education. Recent graduates were also asked how well high school had prepared them for university coursework and for real-world challenges such as competitions, internships, or collaborative academic tasks. These items extended and contextualised the survey results, consistent with the strengths of mixed-methods design (Creswell).

In addition, 12 follow-up interviews of approximately 30 to 40 minutes each were conducted with volunteers, including eight current students and four graduates. These interviews

deepened interpretation of the survey findings by adding detail, context, and explanation to patterns identified in the quantitative responses (Creswell; Palinkas et al.).

Data Analysis

For the quantitative portion, we summarized responses using descriptive statistics and looked for notable patterns. We were particularly interested in gaps between different types of responses – e.g., the percentage of students rating math as highly useful vs. the percentage feeling confident in using math for an unfamiliar problem. We also compared subgroups (there were no major gender differences in this sample given the focus, but we checked for any; none were significant). We treated the sample as a case study of high-achieving Beijing students, not aiming for broad generalizability beyond similar populations, but rather for insight into the magnitude of the skills gap in this context.

Key survey results were tabulated for clarity. Table 1 below highlights a few core indicators that emerged from the data:

Table 1. Key Survey Results

Survey Indicator	Positive Response
“Math is crucial for real-world problems” – agreed (strongly or somewhat)	85%
Confident applying classroom knowledge to new real situations – self-reported “very confident”	25%
Completed an interdisciplinary project in school – answered “yes”	18%
Communication/Teamwork skills are strong – self-rated agree/strongly agree	22%

Note. Positive response indicates the share of respondents who expressed the indicated agreement or experience.

Table 1: Key survey results highlighting belief vs. skill gaps (N = 312). “Positive response” indicates the share of respondents who expressed the indicated agreement or experience.

We can see a striking contrast: an overwhelming majority of these students believe in the importance of math for real-world problem-solving (85% concurred that advanced math is essential in life), yet only one-quarter felt very confident in their own ability to apply what they learned to novel real-world scenarios. Similarly, fewer than one in five had ever participated in a project that crossed disciplinary lines, and only about a fifth rated their soft skills highly.

For the qualitative data, we performed thematic analysis. We coded responses and interview transcripts for themes such as “overconfidence in academic skills,” “difficulty in real-world tasks,” “lack of opportunities,” “test vs. understanding,” and “soft skills deficit.” From this, we identified recurring motifs and representative quotes which will be presented in the Findings section. The qualitative insights are crucial for understanding the why behind the

numbers – they give context to what students feel is missing and how they interpret their own educational experience.

It's worth noting that since this survey targets high-performing students in a specific locale (Beijing), there may be some response bias. These students know they are “good at math” in the conventional sense (indeed many have top Gaokao math scores). Some expressed pride in their achievements but were also surprisingly candid about their limitations. We assured anonymity and framed the survey as a self-reflection on learning, not a test, to encourage honesty. The candor in many responses suggests that even among high achievers, there is an awareness of and frustration with the limitations of their schooling.

In the next section, we present the findings in detail, integrating the quantitative trends with quotes and stories from respondents. This combined evidence paints a clear picture of how Beijing's STEM education, for all its strengths, often leaves students ill-prepared in some key dimensions.

Findings

1. Overconfidence in Math's Utility vs. Underconfidence in Application

One of the clearest findings is the mismatch between students' faith in their academic knowledge and their comfort in using it outside academic settings. As shown in Table 1, 85% of surveyed students agreed that advanced mathematics is “crucial for solving real-world problems.” Many wrote that they had been taught from a young age that math is the “foundation of all sciences” and that doing well in math would enable them to solve complex problems in technology, finance, engineering, and other fields. This reflects what we might call a cultural overconfidence in math's utility: a belief, reinforced by teachers and parents, that excelling in math is both necessary and sufficient for handling real-life challenges.

However, when asked about actual experience or confidence in applying math and science knowledge, the picture changed. Only 25% felt “very confident” that they could apply classroom learning to unfamiliar real-world scenarios. Around 40% were “somewhat confident,” often with caveats, and roughly one-third admitted they were not confident at all. To illustrate, one survey question presented this scenario: “Imagine you are helping to plan a city park, which involves budgeting (math), understanding materials (science), and considering community needs. How confident are you that you could contribute useful analysis or ideas based on what you learned in school?” A high-achieving student with top marks in math responded, “Honestly, I wouldn't know where to start. I can solve equations, but planning a park is so broad – we never did anything like that.” This sentiment, “we never did anything like that,” appeared repeatedly. Students recognise a gulf between textbook problems, even difficult ones, and messy real-world problems.

Qualitative responses added further detail. Many students said their problem-solving practice was almost entirely confined to well-defined contexts: “Find x given this formula,” or “calculate the result of this experiment with perfect data.” When they faced open-ended tasks, they felt lost. As one respondent put it, “In math class, I can tackle difficult integrals or geometry

proofs. But when a question doesn't tell you exactly what to solve for, I freeze. We're not used to figuring out the question itself." This suggests underdeveloped problem-formulation skills, a critical aspect of applying logic in real life. Another student described a national innovation competition: "We had to design a solution for reducing food waste. I thought, I'm good at science, so this should be easy. But I realized I didn't know how to even approach it – it wasn't a single-answer problem. I ended up just doing a lot of internet research. I felt all my physics and chemistry knowledge was sitting in separate boxes that I couldn't connect to this real problem." In the end, the student said their team's project became more of an essay than a scientific solution, and they did not win a prize. The experience made them question what their straight-A transcript really meant in practice.

A small number of students did not see this gap as a failure of education, but as something inevitable. One wrote, "Real life problems are always different. School can't prepare you for everything. At least math makes my brain sharp, so I can learn whatever is needed later." This view echoes a classic defence of rigorous theoretical training: that it develops general intellectual capacity even if it does not teach every practical task directly. There is some basis for this argument. Chinese STEM students have been found to enter higher education with comparatively strong academic and higher-order thinking skills by international standards. However, educational theory also makes clear that transfer from one context to another does not happen automatically. Without deliberate opportunities to connect, abstract, and reapply knowledge, instruction may leave students with strong formal skills but weak practical transfer (Loyalka et al.; Perkins and Salomon).

Our own findings suggest exactly that tension. Even students who believed mathematics had "made their brain sharp" often admitted that they had not actually been taught how to transfer or apply knowledge beyond familiar academic settings. In practice, that responsibility was left for them to figure out later.

The overconfidence seems to stem from the high status of mathematics and exam performance. Students, and often their parents, may assume that because they can solve difficult mathematical problems, they will naturally become effective problem-solvers in any domain. The underconfidence appears when they face tasks without a clearly recognised method or formula. As one university freshman explained after starting an internship, the challenge was not advanced mathematics itself, but having to choose and justify a method under real-world ambiguity.

In sum, the findings indicate that Beijing's STEM students place enormous value on mathematical and scientific knowledge, a valuation reinforced by schools, families, and broader social narratives about STEM success. Yet a clear gap remains in confidence when that knowledge must be used in practical, ill-defined, or unfamiliar situations. In effect, many students have been trained to excel at well-structured puzzles, but not yet to feel secure in solving unstructured problems. One major reason for this appears in the next finding: the relative lack of interdisciplinary and experiential learning opportunities.

2. Lack of Interdisciplinary and Real-World Learning Experiences

The survey results on interdisciplinary projects and real-world problem exposure provided some of the clearest quantitative evidence of the skills gap. Only 18% of respondents said they had completed an interdisciplinary project during high school. Among those who had, the most common examples were science fairs, Model United Nations conferences, or robotics competitions. Importantly, these were usually extracurricular or one-off activities rather than normal parts of classroom instruction. The vast majority, 82%, had never taken part in a formal project requiring them to combine knowledge across subjects or apply classroom learning to a realistic problem beyond textbook exercises.

This pattern is consistent with broader research on Chinese education reform. Although national curriculum reform has long promoted more holistic and quality-oriented learning, implementation has remained constrained by examinations, limited resources, and deeply embedded incentives. More recent STEM literature also argues that soft skills, teamwork, and applied learning have historically not been a major part of Chinese secondary and tertiary STEM education, even though such capacities are increasingly important in both education and the workplace (Dello-Iacovo; Van Horne and Rakedzon).

One student explained in an interview, “We had physics experiments in lab class, but they were just to verify known formulas – nothing open-ended. Every subject stayed in its lane. We never did things like ‘use math and science together to solve X.’ That only maybe happens in some competitions for a few students.” For most respondents, interdisciplinary learning, more common in capstone, project-based, or inquiry-based models, was not part of their ordinary academic experience. Even integration across STEM subjects was limited, and links between STEM and the humanities or social issues were rarer still.

A particularly telling example came from a student who had briefly joined an international exchange programme. She recalled that a U.S. environmental science project forced her to combine physics, economics, and persuasive presentation in ways she had never encountered before. She described it as stressful but more educational than many classes at home. Her account highlights what seems to be missing from many respondents’ Beijing school experiences: projects that require students not only to know content, but to frame a problem, combine methods, communicate clearly, and defend choices in public.

The weakness in soft-skill development was reinforced by the survey results on teamwork and communication. Only about 22% of students felt confident in their communication or teamwork abilities, and fewer than 10% said that presentations or group work were frequent features of their regular classes. As one student put it, “In our school, if you can solve problems fast and correctly on paper, you’re a good student. Nobody cares if you can communicate or work in a team. In fact, working in teams was often discouraged except in PE class, because everything academic was individual.” This matters because it suggests that the weakness is not simply personal shyness or student preference, but something built into the instructional structure itself. That interpretation is also consistent with broader research arguing that Chinese STEM education has underdeveloped soft skills relative to technical skills, and that active-learning frameworks can help close that gap (Van Horne and Rakedzon).

One encouraging pattern, however, was that students who did encounter team-based or interdisciplinary tasks often described them as transformative. A student who had taken part in a national robotics competition said, “We had to build a robot to navigate a maze. It required coding (computer science) and understanding the sensors (physics), and none of us could do it alone. At first, our team struggled – we’re not used to dividing work or brainstorming together. But eventually we figured it out and it was the most memorable learning experience I’ve had. I realized how powerful combining skills could be.” What mattered most in this account was not whether the team won, but that the student became aware of both the limits of individualistic academic habits and the value of combining skills across people and domains.

The lack of real-world context in classroom learning also emerged strongly in the qualitative responses. Students repeatedly noted that mathematics problems were abstract or contrived and that science laboratory activities often served only to confirm known results. Several explicitly said they had rarely been asked how formulas or scientific principles could be used under real conditions. These reflections closely match the broader concern in the literature that reform language has often outpaced classroom transformation, especially where test performance continues to dominate incentives (Dello-Iacovo).

Taken together, these findings strongly reinforce broader calls for more active learning, interdisciplinary tasks, and deliberate soft-skill development in Chinese STEM education. Existing research has already argued that teamwork, communication, collaboration, and application need much more explicit space in Chinese STEM classrooms; the present survey adds ground-level evidence from Beijing that these gaps remain visible even in relatively high-achieving settings (Van Horne and Rakedzon). The picture that emerges is not one of weak students, but of highly capable individual performers who have too rarely been asked to work collaboratively, think across domains, or solve messy real-world problems that do not fit neatly into one subject. That absence of practical and interdisciplinary experience appears to be one of the central drivers of the broader skills gap.

3. Soft Skills Deficits and Logical Reasoning Gaps

Closely related to the findings above, the survey and interviews revealed clear soft-skill deficits that are deeply intertwined with logical reasoning in practice. When this paper refers to “logical application skills”, it does not mean only the ability to solve formal logic problems. It also refers to the ability to reason through complex situations, explain that reasoning clearly, respond to the ideas of others, and revise judgement when circumstances change. In that sense, logical application in real life depends not only on technical competence but also on communication, collaboration, and critical evaluation. This broader understanding is consistent with recent work on Chinese STEM education, which argues that soft skills such as teamwork, communication, and reflective collaboration do not automatically emerge from technical training and instead require deliberate practice and active-learning structures (Van Horne and Rakedzon).

On the soft-skills side, only about one in five students in our survey felt they had strong communication or teamwork abilities. Many admitted that they felt shy, hesitant, or ineffective when articulating ideas aloud. One student wrote, “I can write down a solution, but explaining

how I got it is hard for me – I never had to explain out loud in class.” Another said, “Group discussions were not common, so sometimes I have ideas but I don’t voice them because I’m not sure and not used to speaking up.” These comments matter because even when students possess sound reasoning, a lack of communicative practice can prevent that reasoning from being expressed, tested, or refined through dialogue. Real-world reasoning is rarely a solitary act; it is often sharpened through disagreement, persuasion, and collaborative critique. That interpretation aligns with findings that Chinese STEM classrooms have historically underdeveloped communication and teamwork relative to technical skills, even though those capacities are essential in practice (Van Horne and Rakedzon).

Several respondents also described difficulty with critical evaluation, which sits at the intersection of reasoning and soft skills. One survey item asked how often students were encouraged to question assumptions or challenge an existing solution in class. More than 70% answered “rarely” or “never.” One student elaborated, “In our culture, challenging the teacher or the textbook is not really done. We’re given methods and we practice them. Only maybe in some Olympiad training we think of different approaches, but even then it’s about finding the trick to solve faster, not questioning if the approach is relevant.” This suggests many students are not routinely trained to ask whether a method is appropriate, whether its assumptions hold, or whether an alternative perspective might be stronger. That pattern fits Chen’s argument that critical thinking in Chinese educational settings is shaped not just by cognitive ability, but by sociocultural expectations, classroom norms, and conceptions of what it means to think critically in the first place (Chen).

The qualitative data also suggest that some students perceive a deficit in broad, context-sensitive thinking. One graduate commented, “When I got to university, I met students from other countries. I realized they could approach problems very differently – they had taken courses in philosophy or literature that made them think about society or logic in another way. I felt a bit one-dimensional in comparison, like I only had my math toolset.” This perception connects with a wider policy trend in China: gaokao reform has increasingly been discussed in relation to national demand for STEM talent, and commentary on recent policy adjustments has noted how subject-choice structures are being used to steer more students toward physics and other STEM disciplines. While this may make sense from the standpoint of economic development and university admissions planning, it can also reinforce a narrower educational profile if it is not balanced by broader humanistic and argumentative training (China Policy).

In our interviews, a few students explicitly described what they saw as a “logic gap” in their education. One said he felt highly confident in mathematical logic, such as proofs or programming algorithms, but much less confident when reasoning about everyday decisions or human-centred problems: “I think our education made us good at one kind of logic – deductive problem solving in math – but not so good at logical thinking in general life. For example, analyzing an argument or making a decision with limited data, we didn’t practice those.” This distinction is important. Loyalka et al. found that Chinese STEM students often enter higher education with strong academic skills by international comparison, yet do not show the same gains in critical thinking over time as their counterparts in the United States. That pattern is

consistent with the possibility that students develop powerful formal reasoning in structured settings while receiving fewer opportunities to extend that reasoning into ambiguous, open-ended, or interdisciplinary contexts (Loyalka et al.). Another student pointed to the lack of opportunities for independent judgement: “Everything was decided for us – the steps to solve a problem were outlined. We hardly ever had to choose a method or decide what’s important by ourselves until maybe the Gaokao essay, which is ironically in the Chinese subject, not in STEM.” This remark captures a broader issue running through the survey: many respondents were not saying they lacked intelligence or effort, but that they had rarely been placed in situations requiring method selection, prioritisation, argument evaluation, or open-ended judgement. The issue is not an absence of reasoning ability, but a lack of structured opportunities to practise reasoning under conditions of uncertainty and discussion. That interpretation is consistent with both Chen’s analysis of critical thinking in Chinese contexts and Van Horne and Rakedzon’s emphasis on the importance of active, collaborative learning environments for soft-skill development (Chen; Van Horne and Rakedzon).

A few teachers we spoke to informally, although not part of the formal survey dataset, echoed these concerns. One veteran mathematics teacher said the packed curriculum left little time to explore why formulas mattered or how they could be applied, especially when parents judged teaching mainly by test results. This reflects the systemic pressures identified elsewhere in the paper. Overall, our findings suggest that Beijing’s STEM students often lack practice in the softer side of logical thinking: communicating reasoning, questioning assumptions, evaluating methods, and applying logic in ambiguous contexts (Chen; Loyalka et al.; Van Horne and Rakedzon).

Discussion

The empirical evidence from our Beijing survey reinforces a pattern that has long been recognised in discussions of Chinese education: strong academic performance in STEM coexisting with weaker development in practical, interdisciplinary, and soft-skill domains. This section interprets those findings in a broader context, considers why they matter beyond school, and examines why progress has remained limited despite repeated reform efforts.

Reflection of National Trends and Cultural Context

First, the Beijing findings strongly reflect wider national trends. The dominance of exam-oriented education (*yingshi jiaoyu*) has been criticised in China for decades, and reform efforts since the late 1990s have repeatedly tried to move the system toward *suzhi jiaoyu* or “quality-oriented education” (Dello-Iacovo). Yet our data suggest that even in 2024, students in relatively elite Beijing schools still experience an educational culture heavily shaped by high-stakes testing, content mastery, and narrow definitions of achievement. This interpretation is consistent with earlier accounts from the Harvard Graduate School of Education, which described Chinese schooling as producing high achievers while also crowding out creativity, questioning, and applied learning (Harvard Graduate School of Education).

At the same time, the present findings suggest that the issue is not a simple absence of reasoning ability. Rather, students' reasoning appears context-bound: highly effective within familiar, structured academic tasks, but less secure when problems become open-ended, interdisciplinary, or socially embedded. The problem, then, may not be basic intellectual capacity, but the limited cultivation of transfer, independent judgement, and reflective application (Loyalka et al.).

Beijing may, in fact, represent a best-case version of the national picture. The city is relatively well resourced, hosts many elite schools and universities, and has participated in reform pilots, including new gaokao arrangements and broader attempts to diversify curriculum pathways (OECD; China Policy). If even here our respondents reported very limited interdisciplinary project experience and weak exposure to collaborative, real-world problem-solving, then the pattern may well be stronger in less advantaged contexts. In this sense, Beijing is analytically important not because it is typical in every respect, but because it shows how persistent these gaps remain even where resources, prestige, and reform exposure are comparatively strong.

Culturally, the high status of STEM is reinforced by both state priorities and family expectations. Recent gaokao reforms have been explicitly framed in part around national demand for STEM talent, university subject requirements, and labour-market prospects (China Policy). That broader policy climate is mirrored in families' educational behaviour. Research on parental educational anxiety under the "Double Reduction" era suggests that anxiety around competition, school choice, and children's long-term prospects remains significant, even after attempts to reduce out-of-school academic pressure (Chen et al.). In high-pressure urban environments such as Beijing, this can produce a tunnel effect: mathematics, science, and credential value become so central that broader intellectual development is treated as secondary or deferred.

That helps explain why reform has been difficult to implement in practice. Parents may agree in principle that creativity, communication, and independent thinking matter, but still prioritise exam preparation because the existing system continues to reward it most clearly. Teachers face a similar dilemma. As earlier reform literature shows, curriculum rhetoric can change faster than classroom behaviour when the incentive structure remains dominated by examinations, university placement, and measurable academic outcomes (Dello-Iacovo). Our respondents' comments reflect exactly that tension: they often recognised that something important was missing, but also understood why their schools had not prioritised it.

Real-World Consequences: Beyond the Classroom

The skills gap identified in this study is not merely an educational issue. At the individual level, it can affect students' transition into higher education and employment. Our graduate respondents often described feeling underprepared for self-directed work, teamwork, or open-ended tasks once they left school. That aligns with the international evidence from Loyalka et al., which suggests that strong entry-level academic skills in China do not necessarily translate into continued growth in critical-thinking performance during university. Recent STEM-education research has also emphasised that soft skills such as communication,

collaboration, and teamwork remain underdeveloped relative to technical content in Chinese educational settings, despite their importance in professional life (Van Horne and Rakedzon).

At the societal level, the issue matters because many high-stakes real-world problems are not neatly technical. They involve uncertainty, institutions, incentives, ethics, communication, and the ability to integrate knowledge across domains. That is one reason the historical examples discussed earlier remain relevant. In Japan's late-1980s asset bubble, Shiratsuka argues that excessively optimistic expectations and "euphoria" played a central role in sustaining the boom, illustrating how sophisticated systems and educated elites can still misjudge risk when critical questioning weakens. In China's recent property-market adjustment, the IMF has similarly emphasised deeper structural imbalances, including speculative demand, vacant investor-owned housing, distressed developers, and spillovers into the broader economy (International Monetary Fund). These cases do not prove that educational culture causes crises, but they do support a narrower point: technical skill alone is not enough when decision-makers face complex, system-level problems requiring judgement, scepticism, and interdisciplinary reasoning.

The present findings fit that broader concern. Many respondents described being strong at solving clearly defined problems, but weakly prepared for situations requiring them to choose a method, explain a rationale, or weigh multiple interacting factors. If those habits persist into adult professional life, the result may not be incompetence in a narrow sense, but overconfidence within silos and underpreparedness for ambiguity. That is precisely why practical reasoning, communication, and ethical or social context matter alongside technical mastery.

The findings are not wholly pessimistic. Students who had experienced projects, competitions, exchange settings, or other open-ended tasks usually described them as difficult but valuable. This suggests the problem is not resistance to broader learning, but limited access to it within formal education.

Why Do the Problems Persist Despite Recognition?

China is not short on diagnoses of these problems. For more than two decades, scholars, educators, and policymakers have argued for a more creative, inquiry-driven, and holistic model of education (Harvard Graduate School of Education; Dello-Iacovo). Yet the present findings suggest that implementation remains partial. One major reason is examination inertia. The gaokao continues to function as the central gateway to elite universities and high-status opportunities, which makes schools and families understandably reluctant to de-emphasise whatever most directly improves test outcomes. Even reforms meant to make the system more flexible remain tied to competitive selection logics (China Policy).

A second reason is parental and cultural pressure. Research on parental educational anxiety shows that competition remains intense and that families continue to experience major pressure over their children's schooling trajectories (Chen et al.). In such an environment, exploratory learning can appear risky, especially if its payoff is less immediate or less legible than test scores. Our respondents' remarks about parents viewing projects as distractions from "real studying" reflect exactly this problem: when systems are highly competitive, even actors who support reform in principle may resist it in practice.

A third reason is institutional rigidity. Changing classroom practice at scale requires teacher retraining, altered assessment structures, new materials, and time for teachers to plan differently. Earlier reform literature already identified this gap between policy aspiration and classroom implementation, noting that schools often interpret reform through the lens of existing institutional pressures rather than fundamentally changing pedagogy (Dello-Iacovo). Our survey suggests that this dynamic remains highly relevant. Teachers may recognise the value of deeper reasoning and application, but still feel compelled to prioritise pace, coverage, and measurable results.

A fourth reason is that labour-market and university signals have only partially changed. Although some universities and employers now value research experience, creativity, or communication more openly, the dominant sorting mechanisms still privilege test performance, institutional prestige, and formal credentials. As long as those remain the clearest and most powerful signals, schools have limited incentive to reorganise teaching around less directly rewarded capabilities. This is one reason Beijing's gaokao reforms, while important, have not automatically transformed the logic of classroom life.

Taken together, these factors make the problem look less like a failure of awareness and more like a coordination problem embedded in a high-pressure system.

Urgent Reforms Needed

First, curriculum integration and project-based learning should become a formal rather than peripheral part of schooling. If interdisciplinary projects remain extracurricular, optional, or symbolic, they will continue to be crowded out by exam preparation. Structured, curriculum-embedded projects can create opportunities for students to apply knowledge, confront ambiguity, and build communication and teamwork skills in authentic contexts. This recommendation is consistent with long-standing reform arguments for more applied and holistic learning in China (Dello-Iacovo).

Second, assessment reform is essential. So long as the most important assessments reward primarily individual mastery of academic content, schools will rationally continue to optimise for that outcome. Expanding the role of applied tasks, open-ended reasoning, or project components would not be simple, especially in a system concerned with scale and fairness, but without some change in what counts, incentives are unlikely to shift meaningfully. Even the logic of PISA underscores this point: the assessment was designed to value the application of knowledge in real-life situations, not just the reproduction of learned content (OECD).

Third, teacher training and incentives need to be aligned with the reform agenda. Teachers cannot be expected to facilitate active, interdisciplinary, communication-rich learning if they are trained, evaluated, and time-pressured in ways that reward coverage above all else. Reform therefore requires more than new curriculum documents; it also requires professional development, time, and evaluation systems that treat these practices as central rather than decorative (Dello-Iacovo).

Fourth, soft skills need to be embedded inside STEM instruction rather than treated as separate extras. The evidence from Van Horne and Rakedzon suggests that teamwork and

communication improve when students are actually required to use them in substantive work. That means presentations, peer critique, collaborative problem-solving, and explanation should be normal features of STEM classrooms, not occasional additions. Our own respondents' experiences strongly support this conclusion.

Fifth, students need more meaningful exposure to humanities, ethics, and social context. This is not because STEM should be devalued, but because many of the real problems students will later face sit at the intersection of technical systems and human judgement. A more rounded curriculum can help students move beyond narrow optimisation toward broader evaluation of consequences, assumptions, and competing values. That broader aim also fits calls to reconceive STEM education not only as technical preparation, but as a form of praxis connected to society and lived reality (Ma).

Sixth, extracurricular competitions and applied programmes should be treated as educational laboratories, not just prestige exercises. Several respondents in our study described robotics, science fairs, and exchange experiences as the moments when they first encountered genuine collaboration, uncertainty, and interdisciplinary thinking. Schools could make more deliberate use of such spaces to complement the ordinary curriculum and to expose more students, not just a small self-selected minority, to these modes of learning.

None of these reforms is easy. Yet the cost of inaction may rise over time. China's future challenges are unlikely to be solved by narrow technical proficiency alone; they will require people who can analyse evidence critically, communicate across domains, collaborate under uncertainty, and recognise when a technically elegant solution is socially incomplete. In that sense, strengthening practical reasoning and interdisciplinary competence is part of what real STEM excellence now requires.

Reform also depends on communication. The case for change is strongest when broader skills are framed as practical necessities rather than soft extras: for policymakers, they support national talent goals; for parents, they protect long-term success; and for schools, they make strong academic foundations more transferable, collaborative, and resilient.

Communication also needs to involve students themselves. One of the strengths of the present study is that it shows the gap through students' own voices. The argument becomes harder to dismiss when high-performing Beijing students themselves say they feel underprepared for teamwork, explanation, judgement under uncertainty, or application outside exam settings.

Effective communication should not rely only on fear. Historical cases can serve as warnings, but reform is more persuasive when it is also aspirational. A more well-rounded STEM education would not weaken China's strengths; it would deepen them by adding capacities needed for innovation, leadership, and more resilient judgement in complex situations (Shiratsuka; International Monetary Fund; Ma).

Our study contributes to that communication effort by offering both data and narrative evidence. Findings such as the very low proportion of students reporting interdisciplinary project experience, or the similarly weak confidence in teamwork and practical problem-solving, make the issue tangible. The broader value of such evidence is that it helps bridge the distance between acknowledgement and action. It gives educators, school leaders, and policymakers language they

can use to explain that the issue is not the abandonment of rigor, but the expansion of what rigor should mean in the twenty-first century (OECD; Van Horne and Rakedzon).

Conclusion

Beijing's STEM education exemplifies a central duality in contemporary Chinese schooling. On one hand, it demonstrates an extraordinary capacity to produce high achievement in demanding academic subjects and to cultivate strong formal competence in mathematics and science. On the other, the evidence presented here suggests a continuing underemphasis on practical application, interdisciplinary integration, communication, and collaborative reasoning. Our survey and interview data reveal students who are often justifiably proud of their technical ability, yet noticeably less certain about how to use that ability outside highly structured academic settings. That conclusion is broadly consistent with wider research showing that strong academic skill does not automatically translate into continued growth in critical thinking, transfer, or soft-skill development (Loyalka et al.; Van Horne and Rakedzon).

This gap is not merely an academic concern. It points toward a broader challenge for a society facing increasingly complex economic, technological, and environmental problems. The historical cases discussed in this paper are not proof that schooling alone causes crises, but they do illustrate a narrower and important point: technical sophistication without broader judgement can leave institutions and decision-makers vulnerable when problems become systemic, uncertain, and interdisciplinary. Japan's asset bubble and China's property-sector adjustment both demonstrate that highly educated systems can still fail to question assumptions early enough or to integrate risk across domains (Shiratsuka; International Monetary Fund).

At the same time, the picture is not hopeless. China has already articulated many of the policy ambitions needed for change. The language of quality-oriented education, educational modernization, and a high-quality education system is not absent from official discourse; indeed, it is central to it (Dello-Iacovo; CPC Central Committee and State Council). Beijing, because of its concentration of top schools, policy visibility, and institutional resources, is particularly well positioned to become a testing ground for the next stage of reform. If project-based learning, interdisciplinary study, communication-rich classrooms, and broader definitions of achievement can be normalised there, they could provide a model with wider national influence.

The central challenge, then, is not only identifying what needs to change, but building enough alignment to make the change real. That means persuading parents that broader skills are not a threat to success, persuading schools that rigor can include application and dialogue, and persuading policymakers that innovation goals require more than technical acceleration alone. Practical reasoning, ethical reflection, and the ability to communicate and collaborate are not peripheral to STEM excellence; they are part of its mature form (Ma).

In that sense, the goal of reform is not to diminish Beijing's or China's STEM strengths, but to complete them. The existing foundation of technical knowledge is a major asset; what must be added are the capacities that allow knowledge to travel: practical judgement, interdisciplinary thinking, communication, critique, and adaptability. The transformation required is well captured by the phrase *zhi xing he yi* (知行合一), the unity of knowing and

acting. If that shift can be communicated persuasively and implemented seriously, the true prestige of the system will lie not only in elite test results, but in its ability to produce graduates capable of turning knowledge into wise and effective action (OECD; Ma).

Works Cited

- Beijing Municipal Government (2026) Report on the Work of the Government 2026. Beijing: Beijing Municipal Government. <https://english.beijing.gov.cn/government/reports/>.
- Chen, G., Oubibi, M., Liang, A., and Zhou, Y. (2022) ‘Parents’ educational anxiety under the “Double Reduction” policy based on the family, and students’ personal factors’, *Psychology Research, and Behavior Management*, 15, pp. 2067–2082. <https://doi.org/10.2147/PRBM.S375579>.
- Chen, L. (2017) ‘Understanding critical thinking in Chinese sociocultural contexts: A case study in a Chinese college’, *Thinking Skills, and Creativity*, 24, pp. 140–151. <https://doi.org/10.1016/j.tsc.2017.02.015>.
- China Policy (2024) ‘Gaokao reform reflects focus on STEM’. <https://policycn.com/public/commentaries/gaokao-reform-reflects-focus-on-stem-38425> (China Policy).
- CPC Central Committee, and State Council (2025) Master plan on building China into a leading country in education (2024–2035). Beijing: CPC Central Committee, and State Council. https://english.www.gov.cn/policies/latestreleases/202501/20/content_WS678d85c6c6d0868f4e8eef83.html.
- Creswell, J.W. (2006) ‘Choosing a mixed methods design’, in Creswell, J.W., and Plano Clark, V.L. (eds.) *Designing, and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage, pp. 58–88.
- Dai, D.Y., Gerbino, K.A., and Daley, M.J. (2011) ‘Inquiry-based learning in China: Do teachers practice what they preach,, and why?’, *Frontiers of Education in China*, 6(1), pp. 139–157. <https://doi.org/10.1007/s11516-011-0125-3> ([https://link.springer.com/article/10.1007/s11516-011-0125-3? link](https://link.springer.com/article/10.1007/s11516-011-0125-3?link)).
- Dello-Iacovo, B. (2009) ‘Curriculum reform, and “Quality Education” in China: An overview’, *International Journal of Educational Development*, 29(3), pp. 241–249. <https://doi.org/10.1016/j.ijedudev.2008.02.008> (SCIRP).
- Guo, J., and Qu, X. (2022) ‘Competition in household human capital investments: Strength, motivations, and consequences’, *Journal of Development Economics*, 158, article 102937. <https://doi.org/10.1016/j.jdeveco.2022.102937>.
- Harvard Graduate School of Education (2003) ‘Learning in China’, *Ed. Magazine*, 1 October. <https://www.gse.harvard.edu/ideas/ed-magazine/03/10/learning-china> ([https://www.gse.harvard.edu/ideas/ed-magazine/03/10/learning-china? – HGSE link](https://www.gse.harvard.edu/ideas/ed-magazine/03/10/learning-china?–HGSE+link)).
- International Monetary Fund (2024) People’s Republic of China: Selected Issues. IMF Country Report No. 24/50. Washington, DC: International Monetary Fund. <https://www.elibrary.imf.org/downloadpdf/view/journals/002/2024/050/article-A000-en.pdf> ([https://www.elibrary.imf.org/downloadpdf/view/journals/002/2024/050/article-A000-en.pdf? IMF library](https://www.elibrary.imf.org/downloadpdf/view/journals/002/2024/050/article-A000-en.pdf?IMF+library)).
- Likert, R. (1932) ‘A technique for the measurement of attitudes’, *Archives of Psychology*, 140, pp. 1–55.

- Loyalka, P., Liu, O.L., Li, G., Kardanova, E., Chirikov, I., Hu, S., Yu, N., Ma, L., Guo, F., Beteille, T., Tognatta, N., Gu, L., Ling, G., Federiakin, D., Wang, H., Khanna, S., Bhuradia, A., Shi, Z., and Li, Y. (2021) ‘Skill levels, and gains in university STEM education in China, India, Russia, and the United States’, *Nature Human Behaviour*, 5, pp. 892–904. <https://doi.org/10.1038/s41562-021-01062-3> (<https://www.nature.com/articles/s41562-021-01062-3>? Link to article).
- Ma, Y. (2021) ‘Reconceptualizing STEM education in China as praxis: A curriculum turn’, *Sustainability*, 13(9), article 4961. <https://doi.org/10.3390/su13094961> (<https://www.mdpi.com/2071-1050/13/9/4961>? – link).
- Ministry of Education of the People’s Republic of China (2001) Notice on issuing the Basic Education Curriculum Reform Outline (Trial) [基础教育课程改革纲要(试行)]. Beijing: Ministry of Education of the People’s Republic of China. https://www.moe.gov.cn/srcsite/A26/jcj_kcjcgh/200106/t20010608_167343.html (Ministry of Education China).
- Ministry of Education of the People’s Republic of China (2021) 新高考带来哪些新变化——高考综合改革七周年回眸. Beijing: Ministry of Education of the People’s Republic of China. https://www.moe.gov.cn/jyb_xwfb/s5147/202109/t20210914_562841.html (https://www.moe.gov.cn/srcsite/A26/jcj_kcjcgh/200106/t20010608_167343.html? – ministry of education link).
- OECD (2016) Education in China: A Snapshot. Paris: OECD Publishing. <https://doi.org/10.1787/9789264258490-en>.
- OECD (2019a) PISA 2018 Results (Volume I): What Students Know, and Can Do. Paris: OECD Publishing. <https://doi.org/10.1787/5f07c754-en>.
- OECD (2019b) B-S-J-Z (China) – Country Note – PISA 2018 Results. Paris: OECD. https://www.oecd.org/pisa/publications/PISA2018_CN_QCI.pdf.
- OECD (2019c) PISA 2018 Assessment, and Analytical Framework. Paris: OECD Publishing. <https://doi.org/10.1787/b25efab8-en>.
- OECD (2020) Benchmarking the Performance of China’s Education System. Paris: OECD Publishing. <https://doi.org/10.1787/4ab33702-en> (https://www.oecd.org/en/publications/benchmarking-the-performance-of-china-s-education-system_4ab33702-en.html? – OECD link).
- Palinkas, L.A., Horwitz, S.M., Green, C.A., Wisdom, J.P., Duan, N., and Hoagwood, K. (2015) ‘Purposeful sampling for qualitative data collection, and analysis in mixed method implementation research’, *Administration, and Policy in Mental Health, and Mental Health Services Research*, 42(5), pp. 533–544. <https://doi.org/10.1007/s10488-013-0528-y>.
- Perkins, D.N., and Salomon, G. (1992) ‘Transfer of learning’, in Husén, T., and Postlethwaite, T.N. (eds.) *The International Encyclopedia of Education*. 2nd edn. Oxford: Pergamon Press.
- Shiratsuka, S. (2005) ‘The asset price bubble in Japan in the 1980s: Lessons for financial, and macroeconomic stability’, in *Real Estate Indicators, and Financial Stability*. BIS Papers

- No. 21. Basel: Bank for International Settlements, pp. 42–62.
<https://www.bis.org/publ/bppdf/bispap21e.pdf>.
- The State Council of the People's Republic of China / National Development, and Reform Commission (2021) *The Outline of the 14th Five-Year Plan for National Economic, and Social Development of the People's Republic of China, and the Long-Range Objectives Through the Year 2035*. Beijing: State Council / NDRC.
<https://en.ndrc.gov.cn/policies/202203/P020220315511326748336.pdf>.
- Van Horne, C., and Rakedzon, T. (2024) 'Teamwork made in China: Soft skill development with a side of friendship in the STEM classroom', *Education Sciences*, 14(3), article 248.
<https://doi.org/10.3390/educsci14030248> (<https://www.mdpi.com/2227-7102/14/3/248?> – MDPI link).
- Wang, L., Liu, Q., Du, X., and Liu, J. (2017) 'Chinese mathematics curriculum reform in the 21st century: A review', *Eurasia Journal of Mathematics, Science, and Technology Education*, 13(8), pp. 5311–5326. <https://doi.org/10.12973/eurasia.2017.01005a>
(<https://www.ejmste.com/article/chinese-mathematics-curriculum-reform-in-the-21st-century-a-review-4949?> – Eurasia Journal link).
- Wang, T. (2019) 'Competence for students' future: Curriculum change, and policy redesign in China', *ECNU Review of Education*, 2(2), pp. 234–245.
<https://doi.org/10.1177/2096531119850905>
(<https://journals.sagepub.com/doi/abs/10.1177/2096531119850905?> Sage Journals link).
- Wang, Y., Wang, H., Xue, M., and Qiao, X. (2024) 'Navigating educational uncertainties: Middle-class parental anxieties in China's "Double Reduction" era', *Chinese Education & Society*, 57(3–4), pp. 196–216. <https://doi.org/10.1080/10611932.2024.2441145>
(<https://www.tandfonline.com/doi/full/10.1080/10611932.2024.2441145?> - link).

From Iron and Blood to Empire: The Fall of Liberal Nationalism in Europe

By Daniel Kim

In 1815, following the fall of Napoleon's empire, Klemens von Metternich stood at the helm of Europe's reconstruction, determined to ensure that the tried institutions of monarchs and aristocrats would never again be shaken by the "presumption"¹ of revolution. "Let [monarchs] maintain religious principles in all their purity,"² he wrote, warning that if men dared to swerve from the "sovereign arbiters of their destinies,"³ they would plunge the continent into convulsion. For decades, his system held. But by 1848, revolts again erupted from Palermo to Berlin, and

Metternich was forced to flee Vienna in disguise. The world he had warned against had arrived. In the years that followed, a new generation of leaders—Bismarck, Cavour, and Napoleon III—emerged, taking advantage of the weakening consensus among Europe's Great Powers. The Crimean War shattered the unity of the Concert of Europe in 1853–1856, and, as former allies turned against each other, common diplomatic principles unraveled. Abandoning collective restraint, these realpolitik leaders pursued national ambition through war and calculated alliances, presenting their actions as efforts to achieve patriotic unity. Political success no longer depended on ideological purity or popular mandates. Instead, leaders like Bismarck, Cavour, and Napoleon III rose to power by capitalizing on nationalist zeal to justify authoritarian rule, while abandoning ideological commitments in favor of opportunistic diplomacy.

Post-1848, these leaders succeeded by severing nationalism from its earlier ties to liberalism and capitalizing on nationalistic sentiment instead to legitimize authoritarian rule. Across Europe, reformist movements had splintered apart. "Organized nationalism more or less ceased to operate [...] The few who favoured constitutional reform and national unity began to call themselves 'liberals' [...] but they were unable to agree on any kind of common programme."⁴ Liberal-nationalist ideals had failed to produce stability or reform, and leaders like Bismarck moved quickly to fill the void, redefining nationalism as an initiative that required strong, centralized authority. "Not by speeches and majorities will the great questions of the day be decided—that was the mistake of 1848 and 1849—but by iron and blood,"⁵ he declared to the Prussian Reichstag, rejecting parliamentary debate and condemning liberal-nationalist methods of unification through civic consensus in favor of military strength and state power. Louis-Napoleon adopted a similar approach, defending his authoritarian rule on the premise that liberal reformers had proven ineffective. "Disabused of absurd theories,"⁶ he claimed, "the people have acquired the conviction that the pretended reformers were only dreamers."⁷ He positioned himself not as an ideologue, but as a stabilizer who could restore order and national strength without democratic institutions. By denouncing liberal failure and claiming sole authority to unify the nation, each leader redefined nationalism as a vehicle of authoritarian rule.

Once in power, these leaders moved quickly to entrench state authority, using nationalism to justify centralized control, military buildup, and the repression of dissent. They claimed that only a strong state guided by elites could preserve the nation's integrity. Just three years after his election, Louis-Napoleon "dissolved the Constituent Assembly, imprisoned many of his

opponents [...] and sent the army into the streets of Paris, where 400 people were killed.”⁸ Although elected democratically, he made it clear that unity would be maintained by force, not through public participation. Bismarck shared this view. “In order to take part in public life,” he wrote, “one must be a salaried and dependent servant of the state.”⁹ The state, independent from the people, was the engine of political legitimacy. Cavour, too, pursued national development through technocratic reform, not civic engagement. “He reorganized the treasury, reformed the taxation system, and borrowed heavily from abroad [...] to maintain and equip a large modern army.”¹⁰ Like Bismarck and Napoleon III, he tied national progress to the strength of state institutions. Bureaucracy, military expansion, and economic modernization were not tools that empowered the public; instead, they consolidated control in the hands of the elites.

With domestic control secured, these leaders turned outward, and, taking advantage of the political vacuum left by the collapse of the Vienna system, Bismarck, Cavour, and Napoleon III abandoned ideological rigidity in favor of opportunistic diplomacy. Cavour, recognizing that “Sardinia alone could not vie with Austria for supremacy,”¹¹ turned to Napoleon III for support, not out of shared ideals, but out of strategic necessity. He offered France territorial concessions—the provinces of Nice and Savoy—in exchange for military intervention in Lombardy, despite “the people of Nice, by origin, language, and customs, [being] closer to Piedmont than France.”¹² This agreement rested less on liberal or nationalist principles than on pragmatic coordination to isolate Austria and shift the balance of power. Bismarck operated with the same flexibility. Referring to his manipulation of the Ems Dispatch, he wrote, “Success [...] depends upon the impression which the origination of the war makes [...] it is important that we should be the party attacked,”¹³ . Rather than declare war outright, he engineered a diplomatic crisis that would paint Prussia as the victim, ensuring public support at home and diplomatic neutrality abroad. His strategy abandoned the preset notions of conservative alliance among the Great Powers, seizing instead on a moment of diplomatic instability to unify Germany on Prussia’s terms. Napoleon III likewise relied on strategic foreign engagement. “Louis-Napoleon may have disliked wars, but he had nonetheless involved France in the Crimean War [...] the war against Austria in Lombardy in 1859 [...] expeditions to Syria and Cochin-China.”¹⁴ These campaigns were not ideological; they were designed to project French strength and restore influence lost after 1815. In each case, diplomacy was driven by timing, opportunity, and the pursuit of national advantage, rather than any fixed principle.

By redefining nationalism as a project of state authority rather than popular empowerment, Bismarck, Cavour, and Napoleon III set a precedent that would entirely reshape European politics. In divorcing nationalism from liberalism, they built nations on the strength and control of the elite, rather than civic participation. Constitutional reform and popular sovereignty, once intertwined with nationalist hopes, were sidelined in favor of state-led consolidation. While this approach may have succeeded in redrawing the map of Europe, it also narrowed the meaning of national unity, linking it not to public consensus but to the power and centralized authority held by a few. The conservative domination of Europe had failed, but so too had the possibility of a unified Europe rooted in democratic ideals. Metternich had once called

on monarchs to wield power “only for the maintenance of tranquility [...] menacing only the disturbers of public tranquility,”¹⁵ but, despite his warnings, by the mid-19th century, power itself had become the end, and a new generation of conservative leaders now ruled.

Works Cited

1. Tierney & Scott, *Western Societies*, 216.
2. *Ibid.*, 218.
3. *Ibid.*, 216.
4. Evans, *The Pursuit of Power*, 34.
5. Tierney & Scott, *Western Societies*, 316.
6. Napoleon III, *Bordeaux Address*, 1.
7. *Ibid.*
8. King, *The Judgment of Paris*, 1.
9. Evans, *The Pursuit of Power*, 36.
10. Rich, *The Age of Nationalism and Reform*, 1.
11. *Ibid.*
12. Tierney & Scott, *Western Societies*, 310.
13. *Ibid.*, 318.
14. King, *The Judgment of Paris*, 1.
15. Tierney & Scott, *Western Societies*, 218.

The Science of Sound: How Music Can Support Neurodivergent Individuals

By Hayden Kuo

Abstract

The objective of this paper is to review evidence on how music and music therapy support self-regulation in neurodivergent individuals with autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD), strengthened by the inclusion of two expert educator interviews. Studies exploring the role of music therapy in strengthening self-regulation, identity, and well-being were reviewed in this paper. To support the literature and highlight areas for future research, two case study interviews were conducted with nationally recognized music educators. Results highlight four common themes: 1. neurodiversity-affirming perspectives that value differences as strengths, 2. unique musical skills developed by neurodivergent learners, 3. music as a tool to improve focus, persistence, and organization, and 4. music as a therapeutic practice that lowers stress and fosters belonging. Limitations include the heavier research focus on autism over ADHD, the lack of studies on adolescents, and the small sample of educators interviewed. Overall, the findings suggest that music provides more than artistic growth, it is also a tool for confidence, resilience, and self-regulation in neurodivergent students.

Keywords

Autism Spectrum Disorder (ASD); Attention-Deficit/Hyperactivity Disorder (ADHD); Music Therapy; Self-Regulation; Neurodiversity; Emotional Regulation; Social Connection; Executive Function; Stress Reduction; Adolescents; Inclusion.

1. Introduction

In 2022, the Centers for Disease Control and Prevention reported that the prevalence of autism spectrum disorder (ASD) among 8-year-olds in the United States was 32.2 per 1,000 children, roughly one in every 31. At the same time, a national survey estimated that 7 million U.S. children aged 3–17 years (11.4%) had been diagnosed with attention-deficit/hyperactivity disorder (ADHD), about one million more than in 2016. Together, ASD and ADHD represent some of the most common neurodevelopmental conditions affecting children and adolescents.

A shared challenge for many with these conditions is self-regulation, the ability to manage emotions and behaviors in ways that fit the situation. Emotional regulation includes calming oneself when angry or lifting one's mood when feeling down, while behavioral regulation involves following through on responsibilities even when motivation is low. Over time, these skills allow individuals to stay on track with goals and commitments.

Because self-regulation difficulties are common, therapies often aim to strengthen these skills. Music therapy has gained increasing attention as a promising intervention for this purpose. Defined by the American Music Therapy Association (2005) as the “clinical and evidence-based use of music interventions to accomplish individualized goals within a therapeutic relationship,” music therapy is used to address emotional, cognitive and social needs. For students with

ADHD or ASD, music's structure and emotional effect may provide ways for sustaining attention, reducing anxiety, and strengthening social connection.

However, the effectiveness of music therapy for self-regulation is not yet fully understood. Most existing studies focus on younger children, while adolescents face distinct challenges such as heavier academic pressure and more complex peer relationships. Since adolescence is also a period of rapid brain development and identity formation, understanding how music therapy can strengthen self-regulation at this stage is especially important.

In this paper, I review recent literature on how music therapy supports self-regulation among neurodivergent individuals and present interviews with two nationally recognized music educators to further support this literature review.

2. Overview of Current Literature on Neurodiversity and Music

Recent scholarship has explored how autism research can better reflect the neurodiversity movement. Hunt and Procyshyn (2024), for example, examined how evolutionary psychiatry (EP) and the neurodiversity movement can work together to improve understanding and support for autistic individuals. EP views autism as a natural cognitive variation, while the neurodiversity movement emphasizes acceptance and accommodations. Their review suggests that autism should not be viewed as a disorder but rather as a different way of thinking, with unique abilities that can be valuable in many areas of life. Their argument is that blending EP with the neurodiversity movement offers a fuller picture. It acknowledges challenges but also highlights the value autistic individuals bring when supported properly. Hunt and Procyshyn's conclusions provide background by framing autism as a natural variation rather than a deficit. This perspective supports the idea that music therapy should not aim to "fix" neurodivergent students but instead provide tools that help them build on their strengths, especially when it comes to self-regulation.

While a lot of research on neurodiversity focuses on societal perceptions, more is needed to show how these ideas can change the way research and therapy are done. Xia and colleagues (2024) have taken a similar position, arguing that traditional medical models often focus too much on deficits, labeling neurodivergent individuals as "different from normal", rather than recognizing their differences as part of human diversity. Their research emphasizes the need to shift away from trying to normalize behaviors and instead create more inclusive approaches that value unique strengths and abilities. For music therapy, this perspective encourages us to see it less as a treatment to correct issues and more to build confidence and support personal growth.

Building on this idea of recognizing strengths, recent work has looked directly at the connection between autism and music. Marquez-Garcia and colleagues (2022) explain that past research often focused too narrowly on "rule-based" skills such as memory, pattern recognition, or attention to detail; they argue that this view leaves out the emotional and social strengths. The authors point out that autistic people are not only strong in technical skills but also capable of emotional expression, improvisation, and creativity in music. By including autistic voices, music appears not just as a structured activity with patterns or rules, but to share emotions, build

identity, and connect with others. The authors conclude that autism and musicality should be understood in a broader way that combines scientific findings with lived experiences, creating a fuller picture of how music plays a role in autistic lives. This wider perspective suggests that music therapy can help students with both self-expression and emotional regulation, as well as technical skills.

In contrast, Abram (2014) placed greater emphasis on the practical application of music therapy. He studied music therapy with children who had autism and complex communication needs. Results demonstrated that music therapy could motivate children, strengthen communication, and improve motor and social skills. At the same time, therapists noted difficulties in ensuring these skills carried over into daily life and in telling the difference between genuine communication from simple repetition. The author concluded that music therapy has real potential, but applying it consistently across different settings requires more research. These findings connect back to self-regulation, since motivation and attention are critical skills for academic and social success. To explore these themes further, I conducted semi-structured interviews and later discussed how these insights supported the conclusions described in the literature review.

3. Methods

This paper presents two semi-structured interviews that were designed to further explore the themes discussed in the literature review above.

3.1 Educator Interviews

Using a qualitative case study design, I interviewed two neurodivergent music educators from Lexington Public Schools. Both are respected by students and colleagues and were named quarterfinalists for the 2025 GRAMMY Music Educator Award, a national recognition of excellence in music teaching. Their achievements and perspectives as neurodivergent educators offer valuable insight for interpreting this study's findings.

Mx. Rachel Jayson. Jayson is a Boston-based violist, conductor, and educator active in both classical and contemporary music. They perform with Walter Sickert & the Army of Broken Toys, the New England Film Orchestra (co-founder), Tapestry, and Zili Misik. As the first African American in Lexington's performing arts department, they have led student orchestras to Symphony Hall, Lincoln Center, and to the 2025 National Grand Champion title at the Orchestra Cup. Jayson also teaches innovative courses such as Mixtape Anatomy, was named a 2024 Diverse String Teacher Fellow, and holds degrees from the Boston Conservatory.

Ms. Jessica Billings-White. Billings-White is a string specialist serving Lexington's elementary, middle, and high schools. She holds advanced degrees in Cello Performance and Music Education from the Boston Conservatory at Berklee and McGill University. In addition to directing programs across districts and teaching pedagogy at the Longy School of Music, she has held leadership roles with MA-ASTA, including President. Earlier, she directed string programs in Ipswich and began her career in New Bedford Public Schools.

3.2 Procedure

I first reached out to the two participants via email to describe the purpose of the interviews and what their participation would look like. After obtaining consent that the participants were comfortable, I scheduled one virtual (questions were sent via email and responses collected via email) and one thirty-minute in-person interview. The interview focused on four areas, including personal journey with music, how music supported challenges in life and acceptance, observed connections between music and self-regulation, and ways music functioned as therapy for students. Before analyzing their responses, I selected four themes based on the interview questions, which were theoretically motivated by past research reviewed earlier, including neurodiversity-affirming perspectives, unique musical strengths and skills, music as a tool to improve skills and behaviors, and music as a therapeutic practice.

3.3 Ethics and Consent

Participants consented to be identified by name and professional role. Only publicly available professional details are included; no private personal information (e.g., contact details, home address, medical data) is reported. The project followed the ethical principles of the Declaration of Helsinki.

4. Results

The findings from the literature and the interviews are integrated below according to the four themes identified in the analysis.

4.1 Neurodiversity-Affirming Perspectives

A common theme across both the literature and interviews was the shift away from deficit-based views on ADHD and autism towards more inclusive perspectives. Research increasingly presents these conditions as natural variations in how people think and learn rather than disorders that must be corrected. This perspective emphasizes acceptance and change, encouraging educational approaches that focus on belonging and support (Hunt and Procyshyn, 2024; Xia et al., 2024).

The educators echoed this idea. Jayson described viewing students not as needing to be “fixed” but as individuals whose differences should be recognized as strengths. “At a time when I felt small, trapped, and without control, I could escape to playing piano alone or music with others. I could see my own progress, and that reminded me I was able to do big, hard things slowly, little by little.” Billings-White emphasized that many challenges students face are invisible and that creating a safe, supportive environment allows them to thrive. She reflected, “I have made it my mission never to give up on a student or judge them at first glance. Life has shown me that we can never truly know what challenges a student has faced before meeting us.” Both the studies and interviews reach a similar conclusion that music works best when it builds confidence and inclusion, not when it tries to ‘fix’ students.

4.2 Unique Musical Strengths and Skills

Research has found strengths such as strong auditory memory, heightened sensitivity to tone and rhythm, and close attention to detail in neurodivergent learners (Marquez-Garcia et al., 2022). These skills are identified as different ways of learning that develop when traditional approaches, like note reading, are difficult. Additionally, these strengths should not be seen as side effects, but as abilities that can be valued on their own.

Both educators shared how students' unique learning profiles shape musical strengths. Billings-White reflected on her own experiences with dyslexia and ADHD, noting, "Because I struggled with note reading for so long, I developed strong listening skills [...] I had to focus intently on tone, intonation, phrasing, and nuance, carefully emulating and memorizing what I heard in order to play it back accurately." She explained that this reliance on auditory memorization through listening became a powerful skill, one she now recognizes in many of her students. Jayson described how sensory experience supports focus: "With my own neurodivergence, it helps me focus, taking the noise and chatter and funneling it into one place. The physical connection and the concentrated gesture like the vibrations of the instrument or bass rattling from the is very regulating."

Both the literature and interviews highlight music as a way for neurodivergent students to not only compensate for challenges but also thrive through their distinct abilities.

4.3 Music as a Tool to Improve Skills and Behaviors

Music was often described as being able to support self-regulation, as learning an instrument requires practice, organization, and attention. Research shows that these habits can be transferred to school and daily life (Abram, 2014).

The educators shared that these habits often are applied outside of music. Jayson emphasized how learning a long piece requires breaking it down into smaller, manageable parts, a process that strengthens planning and organization: "For ADHD especially, music scaffolds big tasks. Learning a piece has inherent structure; you can break it down sequentially, thematically, or tonally. Compare that to something like writing a book music gives clearer pathways, and that helps with executive functioning." Billings-White explained that practicing music in intentional sections helps students build persistence and coping strategies: "One of the most valuable skills I gained from studying music is the habit of writing everything down [...] That skill, simple as it is, has become a cornerstone of how I navigate both my personal and professional life."

Both sources show that music isn't just a form of artistic expression, but a tool that helps students develop self-regulation while strengthening skills like planning, focusing, and staying organized. These benefits support their success in school and in their daily lives.

4.4 Music as a Therapeutic Practice

Studies show that music can lower stress levels and encourage social connection (Abram, 2014; Marquez-Garcia et al., 2022). In clinical settings, music therapy is provided by certified music therapists and is designed to help people achieve individualized goals.

In these interviews, however, the educators described experiences from their classrooms where music acted in a therapeutic way, helping students feel calm and supported even though it was not formal therapy. Billings-White explained that rehearsal spaces often serve as therapeutic environments: “Many students have shared that my class feels like a ‘low-stress zone,’ a space where they can relax, play, and find relief from the busyness of life.” Jayson highlighted the role of ensemble music-making in helping students who might otherwise feel isolated develop a sense of belonging: “When our orchestra director passed away [...] we talked quietly, many of us cried, and then we played. That act of grieving through sound together was the highest form of therapy.”

Their perspectives align with research showing that music contributes not only to technical growth but also to emotional well-being and social connection. These combined results reveal how research and practice work together in shaping self-regulation.

5. Discussion

Both the literature and interviews described music as a tool for self-regulation, but they focused on different parts. Research often focused on technical and cognitive benefits such as attention, memory, and motivation, while educators highlighted emotional well-being, stress reduction, and belonging. These perspectives suggest that music therapy should be valued both for skill development and for creating supportive environments.

Another key finding is that transferable skills can be applied outside the practice room. While research identified persistence and organization, educators provided real examples of students applying these strategies in academics and daily life. This strengthens the argument for music as an intervention for broader self-regulation.

6. Limitations

Although this paper integrates research with educator experiences, there are important limitations.

Most of the studies I reviewed focus more on autism than ADHD. Because of this, the conclusions do not always give equal weight to the experiences of students with ADHD. Future research should look more directly at ADHD, either in separate studies or in side-by-side comparisons with autism. For example, ADHD tends to be diagnosed as hyperactivity in many individuals, meaning that music therapy could serve as a grounding tool for them. The symptoms and traits presented in autism are not always the same as ADHD, and music therapy may be a valuable tool. Similarly, another tool with the same impact but different outcomes based on these nuances.

Additionally, many of the existing studies focus on younger children instead of adolescents. Middle and high school students face different challenges, like heavier school workloads and more complex friendships, so results from younger groups may not fully apply to them. These challenges occur alongside major social and neurological changes, meaning adolescent responses to music therapy could differ significantly from those of younger children.

More studies on adolescents are needed to better understand how music therapy supports self-regulation during this stage of life. For example, one study could explore how music therapy before important assessments improves adolescents' focus and final grades.

Finally, the educator interviews add valuable perspectives but are limited in number. Only two educators from one school district were interviewed. While both are highly respected, their experiences may not reflect the full range of classrooms or students. The purpose of these interviews was to supplement a literature review on this topic, but future research aims to conduct interviews with a large sample of educators in schools. For example, a five-minute classroom routine where one group of students starts class with a music-based activity like quiet listening while another group begins with silent reading could be conducted. Researchers could then compare how students in each group report their focus, stress level, and also look at their class participation and grades over time to find how students may be able to benefit from music.

7. Conclusion

Music therapy and music education show strong potential for helping neurodivergent students, especially when it comes to self-regulation. Research points to several ways music makes a difference, like how it affirms neurodiverse identities, highlights unique strengths like listening and memorization, builds transferable skills such as persistence and organization, and provides a source of stress relief and social connection.

The educators' perspectives supported these ideas and gave them real-life context. Their experiences showed how music classrooms can become safe spaces for acceptance, how students' learning differences can turn into musical strengths, and how the discipline of music practice carries into other parts of life. They also confirmed the role of music in lowering stress and creating belonging through ensemble playing.

Together, the literature and educator insights suggest that music is more than just an art form for neurodivergent individuals, but also a tool for growth, confidence, and resilience. While more research is needed, especially for older students and ADHD populations, the evidence so far makes a strong case that music can play a key role in supporting both learning and well-being.

Works Cited

- Abram, K. Exploring the Impact of Music Therapy on Children with Complex Communication Needs and Autism Spectrum Disorders: A Focus Group Study. Undergraduate honors thesis, Ohio University, 2014. OhioLINK Electronic Theses and Dissertations Center, http://rave.ohiolink.edu/etdc/view?acc_num=ouhonors1399667716.
- American Music Therapy Association. "What Is Music Therapy?" 2005, <https://www.musictherapy.org/about/musictherapy/>.
- Centers for Disease Control and Prevention. "Data and Statistics on ADHD." CDC, 2024, <https://www.cdc.gov/adhd/data/index.html>.
- Hunt, A. D., and T. L. Procyshyn. "Changing Perspectives on Autism: Overlapping Contributions of Evolutionary Psychiatry and the Neurodiversity Movement." *Autism Research*, vol. 17, no. 3, 2024, pp. 459–466. <https://doi.org/10.1002/aur.3078>.
- Marquez-Garcia, A. V., et al. "Music Therapy in Autism Spectrum Disorder: A Systematic Review." *Review Journal of Autism and Developmental Disorders*, vol. 9, no. 1, 2022, pp. 91–107. <https://doi.org/10.1007/s40489-021-00246-x>.
- Xia, Y., P. Wang, and J. Vincent. "Why We Need Neurodiversity in Brain and Behavioral Sciences." *Brain-X*, vol. 2, no. 2, 2024, pp. 1–7. <https://doi.org/10.1002/brx2.70>.

The Impact of Artificial Intelligence and Disclosure in Advertisements on High School Students By Ketan Shah

Abstract

The increased use of AI-generated content creation has greatly expanded in the communications industry. However, early indications show that some target audiences may have differing opinions about the authenticity of AI generation on content, especially visuals (Kirk, Givi, 2025; Desai, Hamsa, 2025; Zhang, Hur, 2025). The purpose of this study was to evaluate two types of advertisements in order to assess their effectiveness in generating positive product perceptions among young adults. The study's survey used a pre- and post-exposure design to compare product perceptions before and after viewing advertising samples labeled as AI generated and not labeled and human produced. Respondents' assessments of product perceptions were categorized in three areas: perceived authenticity, desire to purchase, and willingness to pay a certain amount. The results of this study indicated that an AI-labeled advertisement generates more positive product perceptions than a human-produced, unlabeled advertisement. These results can be used by marketing executives to develop communications that effectively drive purchases and foster authentic product perceptions among high school students. After conducting this experiment, it was discovered that AI-labeled messages are more effective in eliciting more positive product perceptions from high school students.

Keywords advertising, marketing content creation, message framing, purchase intent, inclination, appeal, high school students, AI-authorship, artificial intelligence, perceptions, label

1. Introduction

With the rapid growth in AI use (Crane, Green, and Soto, 2025), many businesses are beginning to use this new technology to create marketing communications quickly and efficiently. By making advertisements more convincing and appealing to consumers, businesses can maximize revenue and profit (Wrexham University, 2024). This study is important, especially due to the lack of previous research comparing AI labeled and produced advertisements with human produced, but unlabeled, messages. My research question is: To what extent does AI-authorship labeling in advertising influence prospective buyers' perceptions of a product compared to unlabeled advertisements?

2. Literature Review

2.1. The AI-authorship effect: Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications (Kirk, Colleen P, and Julian Givi, 2025)

This article provides insight into how disclosing AI as an author affects consumer responses to advertisements and other forms of marketing. This experiment was conducted by

showing a group of participants the same advertisement, and these people were told if the article was either written by AI or a human. After viewing this ad, participants' perceptions of the brand were evaluated through questions regarding authenticity, trust, and loyalty to the brand. Overall, the AI written and labeled messages were seen as less authentic and feelings of moral disgust emerged after reading these articles. While my research has some similarities, there is a major gap because prior studies compare AI-labeled to human-labeled advertisements, whereas my experiment is not labeling the human advertisement, so the participants will not know that they are looking at a human-produced message.

2.2. The Impact of Generative AI Images on Consumer Attitudes in Advertising (Zhang, Hur, 2025)

This research consisted of numerous other individual studies where participants reacted to advertisements either made by AI or a human. One set of participants viewed unlabeled human and AI written advertisements, whereas another set viewing labeled AI and human written advertisements. After viewing, the consumers' attitudes, trust towards the company, and purchase intention were evaluated. This article found that when consumers were not aware of the image source, the attitudes between AI and human did not significantly differ, but a greater contrast was noticed when the source was disclosed to the participant. Once again, despite the fact that this article was helpful in identifying the impact of AI in advertising, it did not attempt to pair an AI-labeled advertisement with a human-unlabeled advertisement, providing an essential opportunity for my research.

2.3. Evaluating the impact of generative AI tools on consumer trust and brand perception: An empirical validation (Desai, Hamsa, 2025)

This article studies how marketing content created by AI impacts consumer trust and brand perception by comparing AI-made advertisements with human-made ones. Participants were shown two advertisements for identical products, one made by artificial intelligence and the other made by a human. This study found that AI-generated content tends to reduce consumer trust and is seen as less authentic and raises privacy concerns in comparison to human-made advertisements. The study also found that what people think about these advertisements affects how they see the brand in the future. Even though this article shows the difference in impact from an AI- vs human-made advertisement, it does not directly pair an AI-labeled advertisement with an unlabeled human advertisement, leaving a key gap that my research is designed to address.

3. Question and Hypothesis

3.1. Research Question

It is crucial for businesses to find effective ways to implement new artificial intelligence technology in the development of marketing content. The research question addressed is: To

what extent does AI-authorship labeling in advertising influence prospective buyers' perceptions of a product compared to unlabeled advertisements?

3.2. Hypothesis

Based on prior research indicating a connection between AI-labeled content and lower authenticity and consumer trust (Kirk, Givi, 2025; Desai, Hamsa, 2025; Zhang, Hur, 2025), products advertised with AI-labeled messages will be perceived as lower in authenticity than non-AI messaging and consumers will be less inclined to purchase the product compared to those advertised with unlabeled human messages. I hypothesize that a marketing communication that is produced by novel artificial intelligence technology and is transparent about it will develop a higher level of negative viewer perceptions than similar messaging produced by traditional human creativity.

4. Methodology

4.1. Sample: GenZ men and women

The survey utilized a convenience sample of 50 high school students in Illinois, New Jersey, California, and Michigan.

4.2. Experiment Method

This experiment tested the impact of the utilization and disclosure of artificial intelligence in marketing on consumer perceptions of a hypothetical product. The data collection tool was a Google Forms survey, which was sent out to about 80 people, and 50 responses were collected, making the response rate about 6 responses per 10 requests. In this survey, participants were initially presented with general questions to assess a baseline knowledge and familiarity regarding the topic of the advertisement, energy drinks. After the initial assessment, participants were presented with the initial exposure message, which was unlabeled and was written by a human. The perceptions of the participants were then assessed using the three indications identified below. Next, participants were presented with the labeled advertisement written by AI and subsequently asked the same post-exposure questions.

The consumer perceptions were assessed in three ways. One assessment consisted of asking respondents to rate their perceived authenticity of the product on a scale from 0 to 5, 5 being the most. This assessment was only applicable, and thus only asked, after exposure to a given advertisement. The following two assessments included questions given both before and after exposure to the AI and human made advertisements, in order to see a direct comparison and effect of the advertisement. One was on the same 0 to 5 scale regarding the desire to purchase the product. Before exposure, this question was framed in a way that asked for the respondent's desire to try or purchase a new energy drink brand. After exposure, the question directly asked the respondent for their inclination to purchase after seeing the advertisement. The third and final

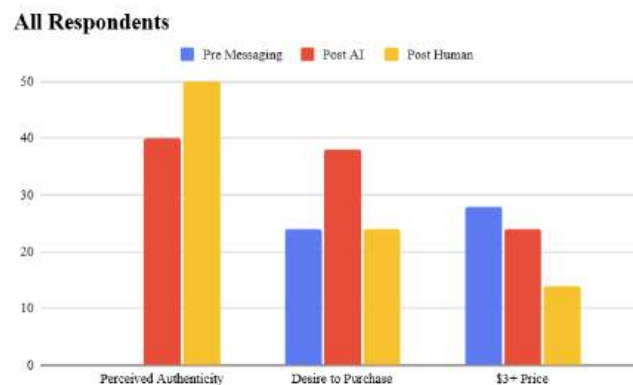
assessment asked approximately how much the respondent would be willing to pay for one unit of the product, and gave ranges such as \$1.01-2 and \$4.01-5 to choose from.

Demographic information on gender was also collected. This information was vital to the survey because it helped understand how donation intent varied by gender.

This experiment was a direct comparison of each subject's change in product perceptions based on the contrast between being exposed to an advertisement labeled with AI-authorship and one not labeled at all and written by a human. I used statistical analysis of the data, which tests the hypothesis that products advertised with a label saying it was written by AI will have worse consumer perceptions.

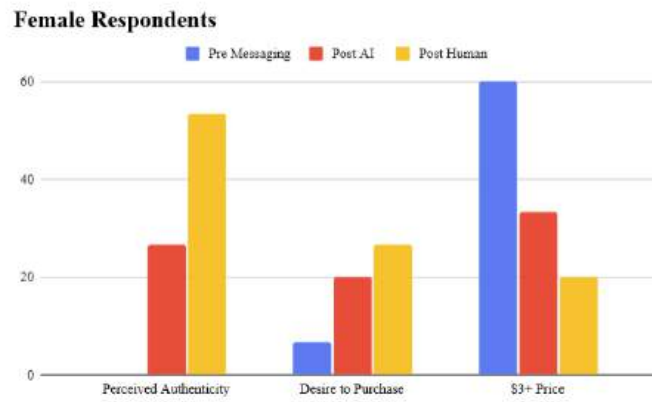
5. Results

5.1. Results from All Respondents



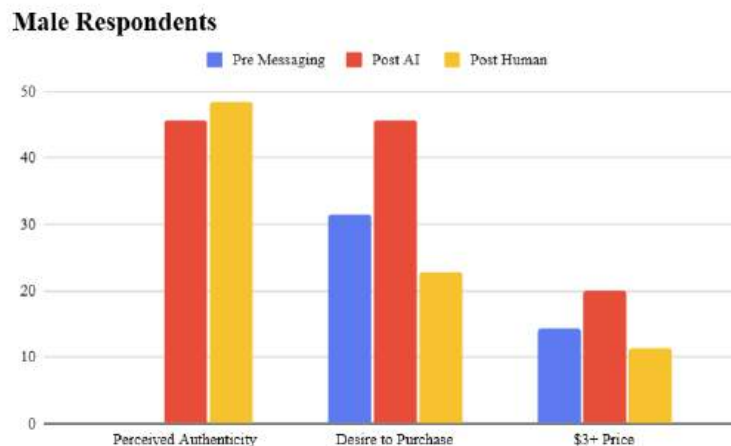
This graph represents all 50 respondents, showing the percentages of respondents who put a 3 or above for the first two metrics on the survey and \$3.01-4 or higher for the price assessment. After collecting the data and performing statistical analysis, each bar peaked in a different assessment regarding consumer perceptions. Overall, the authenticity of the energy drink product was perceived as higher after seeing the post-human advertisement, but with this came a lower desire to purchase the product as well as spend three or more dollars on the product. While the AI-labeled message was seen as less authentic, it was ultimately seen as a more convincing advertisement to purchase and spend more money on. The blue bar in this graph is helpful to realize that the AI-labeled message, not the human-unlabeled message, led to more positive results when comparing initial and final metrics. Given that these three assessments were indications of viewed product perceptions, this means that, overall, the first message, the AI-labeled one, was the more effective advertisement in generating positive perceptions.

5.2. Results from Female Respondents



This graph represents all 15 female respondents, showing the percentages of respondents who put a 3 or above for the first two metrics on the survey and \$3.01-4 or higher for the price assessment. After collecting the data and performing statistical analysis, the human-unlabeled message was once again viewed by the females as more authentic than the AI-labeled one. This message also led to a greater desire to purchase the energy drink product, but the AI-labeled message was more influential to sell a more expensive product. Looking at the blue bar, it is clear that the desire to purchase was higher than initial levels for both advertisements, especially for the human-unlabeled advertisement. However, the willingness to pay \$3 or more for this product significantly dropped after viewing both advertisements, indicating that the female respondents were more willing to pay for this energy drink product regardless of the message, but at a lower price than initially desired. Given that these three assessments were indications of viewed product perceptions, this means that, overall, the second message, the human-unlabeled one, was the more effective advertisement in generating positive perceptions.

5.3. Results from Male Respondents



This graph represents all 35 male respondents, showing the percentages of respondents who put a 3 or above for the first two metrics on the survey and \$3.01-4 or higher for the price assessment. The perceived authenticity between the two advertisements was relatively similar, but the human-unlabeled message ultimately was perceived as higher in authenticity. However, the AI-labeled message produced a higher desire to purchase and willingness to pay more money for one unit of the energy drink product. Furthermore, looking at the blue bar, after the AI exposure, the desire to purchase and willingness to pay \$3+ was higher than pre-messaging levels, indicating that the AI-labeled message exceeded initial expectations. On the other hand, after viewing the human produced message, this desire and willingness dropped below pre-messaging levels. Given that these three assessments were indications of viewed product perceptions, this means that, overall, the first message, the AI-labeled one, was the more successful advertisement in generating positive perceptions.

6. Takeaways and Discussion

6.1. Takeaways

Analysis of the data in this experiment contradicts the hypothesis, which was that advertisements produced and labeled by artificial intelligence will develop a higher level of negative viewer perceptions compared to a similar unlabeled advertisement produced by a human. This experiment instead indicated that human-unlabeled advertisements developed a higher level of negative viewer perceptions, as most respondents felt a lower desire to purchase and spend more money on one unit of the product despite greater sentiments of authenticity. Overall, there was an increase in product perceptions in terms of purchase intent and desired price amount after viewing the AI-labeled message. This finding was mostly consistent across all demographics, however female respondents were more sympathetic to the human-unlabeled message, as they were the only group to have a greater purchase desire after viewing the human advertisement compared to the AI-labeled message. The male respondents data reflected similar patterns as the overall results.

6.2. Discussion and Directions For Future Research

One strength of this study was the use of a within-subjects design, where all respondents were exposed to the same experimental conditions - in this case, both the AI-labeled and human-unlabeled advertisements. This allowed for a more direct comparison of perception before and after the different exposures, which reduces the potential for error that could occur if different respondents were exposed to different advertisements. Another strength of this study was that using energy drinks as the subject of the advertisement shown increased familiarity for the respondents. As high school students who frequently use energy drinks, this topic helped make the survey more relatable to the respondents and likely helped elicit more genuine responses. However, a limitation this study faces is that the order of the advertisements in the study was consistent for all respondents. Having some surveys that start with the AI-labeled

advertisement and others that begin with the human-unlabeled advertisement could lead to differing results due to an order effect. Another potential limitation this study faces is that it only used one product subject. The way in which consumers respond to AI vs human advertisements could be different across differing products, which is something that can be explored in a future study.

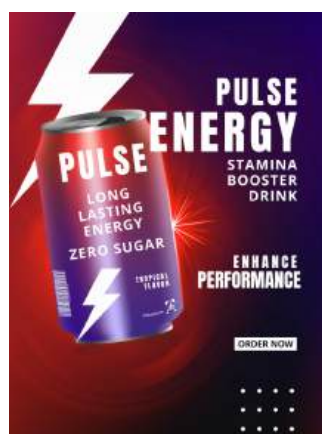
Despite this, it is clear that making messages that are unlabeled and solely made by humans may not be the best method in obtaining product purchases and high product perceptions. Thus, businesses should take this into consideration when trying to obtain high product perceptions from a younger generation. However, due to limitations, even though the AI-labeling strategy is effective, it can potentially not be effective for all product categories. As a result, there is room for further research to explore multiple product types as well as by changing the order of exposure in order to find the best ways to advertise successfully while also coexisting with the powerful tool of AI.

7. Appendix

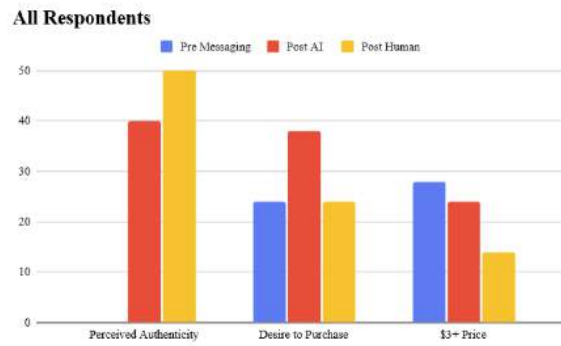
7.1. Appendix A: Initial Exposure Message - AI-Labeled



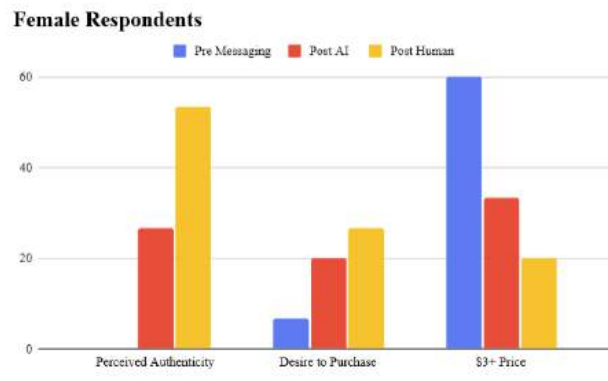
7.2. Appendix B: Second Exposure Message



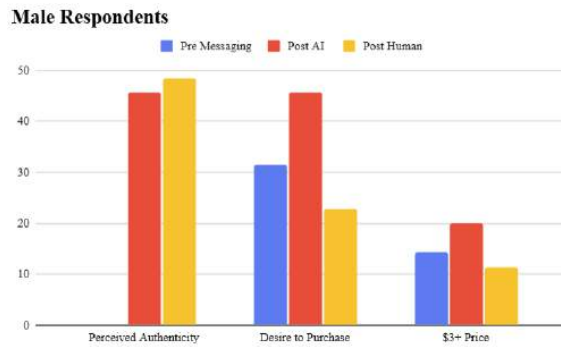
7.3. Appendix C: Graph of Results From All Respondents



7.4. Appendix D: Graph of Results From All Female Respondents



7.5. Appendix E: Graph of Results From All Male Respondents



Works Cited

- Crane, Leland, et al. "Measuring AI Uptake in the Workplace." The Fed - Measuring AI Uptake in the Workplace, 2 May 2025, www.federalreserve.gov/econres/notes/feds-notes/measuring-ai-uptake-in-the-workplace-20240205.html.
- Desai, Sheetal, and Sajeesh Hamsa. "Evaluating The Impact Of Generative AI Tools On Consumer Trust And Brand Perception: An Empirical Validation." Allana Management Journal Of Research, vol. 15, no. 2, 25 Dec. 2025.
- Kirk, Colleen P, and Julian Givi. "The AI-authorship effect: Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications." Journal of Business Research, vol. 186, Jan. 2025.
- Nancholas, Ben. "The Effects of Advertising on Consumer Behaviour." North Wales Management School - Wrexham University, Wrexham University, 19 Mar. 2024, online.wrexham.ac.uk/the-effects-of-advertising-on-consumer-behaviour/.
- Shah, Ketan. "Methods of Advertisement." Google Forms, Google, 10 Jan. 2026, forms.gle/B7wfQNanX6v3xSLf6.
- Zhang, Lei, and Chung Hur. "The Impact of Generative AI Images on Consumer Attitudes in Advertising." MDPI, Multidisciplinary Digital Publishing Institute, 16 Oct. 2025, www.mdpi.com/2076-3387/15/10/395.

AI Classification of Collision Likelihood in Satellite Orbit Paths By Prayag Barath

Abstract

A rapid increase in orbiting satellites and space debris has rendered the prediction of collision likelihood for these materials increasingly difficult. Previously existing models for orbital propagation, such as SGP4, commonly struggle with accuracy due to unmodeled variables such as solar activity and face issues of scalability in application to larger datasets. This project seeks to present a machine learning based approach to amending such issues through the use of Random Forest methodology applied to orbital data from the European Space Agency. Following steps of extensive data cleaning, time series restructuring and normalization, multiple Random Forest models, including a varied number of estimators, were evaluated. Results demonstrated that the eccentricity, inclination, and cross sectional area of the chaser object proved most influential in increasing the prediction accuracy for satellite collision. Ultimately, this project yielded a model with the potential to automate collision avoidance in the future.

Introduction

Human activity in the orbital environment of Earth has dramatically increased over past decades. Recent estimates provide that more than 10,000 active satellites find themselves alongside over 100 million pieces of space debris, the likes of which include dysfunctional satellites, launch canisters and past collision material (European Space Agency, 2026). Such objects travel at unexpectedly high speeds, exceeding 7 to 8 kilometers/second, proving capable of immense damage upon impact (Drolshagen, 2008; Smirnov et al, 2015). As humans continue to launch satellites into low earth orbit, the risk of orbital collisions grows steadily. Earth's orbital environment is divided into 3 categories, separated by altitude. The first of these is the Low Earth Orbit (LEO), extending from 160 kilometers to 2000 kilometers above the surface of the Earth. This region is characterized by orbital periods ranging from 90 to 120 minutes (Hiriart, 2009; Liou, 2013). Consecutively following is the Medium Earth Orbit (MEO), extending from 2000 kilometers to 35,786 kilometers and yielding an orbital period of several hours (Hiriart, 2009). Finally is the Geostationary Orbit (GEO), resting at around 35,786 kilometers above the Earth's surface with an orbital period of approximately 24 hours (Hiriart, 2009). Due to its higher atmospheric drag and its heavy usage by wide arrays of human technology, LEO is most susceptible to higher debris density, and thus experiences a significantly higher rate of orbital collisions (Vatalaro et al, 2002).

An important step in risk mitigation is the development of robust frameworks for collision prediction. Among popular existing propagation models exists the Simplified General Perturbations 4 (SGP4) Model, a set of five mathematical structures, which predicts satellite positioning using two-line element (TLE) datasets; SGP4 accounts for factors including the Earth's atmospheric drag and curvature (Vallado and Crawford, 2006). While SGP4 is effective for short term propagation, its accuracy degrades with larger data sets of orbiting objects and

inability to account for variables including solar activity and atmospheric density (Vallado, David A., et al, 2006).

This project seeks to create a broadly applicable model for the prediction of satellite collision likelihood given a set of flight path influencing factors through the use of a Random Forest machine learning model (Rigatti, 2017; Breiman, 2001). A random forest is an ensemble learning method which describes the plotting of a final outcome based on the values of independent decision trees (value decided by a binary split), and is particularly useful when applied to tabular data (Kavlakoglu, 2025). This method allowed me to better understand the impact of certain conditions on a result, or specific to this project, the impact of distinct atmospheric variables on the orbital path and collision likelihood of a satellite. Within the broader implementation of the Random Forest exist subconcepts of matrix multiplication and the Pearson correlation coefficient, which would enable me to test and compare artificially created data sets against a validation set (Breiman, 2001).

Methods

The goal of this project was to create a robust model for the prediction of satellite orbit collisions. This was accomplished primarily through the use of a Random Forest model. Preceding the training of the model, emphasis was placed on data organization and preprocessing steps seeing as uniformity of data is essential to reliable machine learning outcomes. The data used for this project was the European Space Agency (ESA) Collision Avoidance Challenge dataset, specifically the training dataset provided (European Space Agency, 2026).

Data cleaning involved multiple steps, the first among these being handling time series organization. This ensured that each column contained data associated with its unique collision event, each of which was assigned a corresponding event ID. The cleaning process was accomplished through restructuring of the dataset so that orbital variables were grouped through their time indices. Following time series reorganization was the removal of the categorical column, "c_object_type", (which did not contribute significantly to collision predictions) which would disrupt later examination.

Following cleaning, exploratory analysis was conducted using single-feature and double-feature random forest methods, allowing an understanding of both individual variable as well as inter-variable coexisting influences. Variables were subsequently normalized to confirm consistency of outcomes and to avoid biasing the model towards higher-value variables. Collision likelihood labels were assigned through discrete bin restrictions associated with continuous risk values: a risk value from negative infinity to -15 constituted a high risk, -15 to 5 a medium risk, -5 to 5 a low risk and 5 to infinity a very low risk. These values were identified based on visual inspection of the risk histogram, Figure 1, created earlier in the process. For the normalized dataset, the separations were as follows: a risk value from negative infinity to 0.25 constituted a high risk, 0.25 to 0.5 a medium risk, 0.5 to 0.75 a low risk and 0.75 to infinity a very low risk.

The model was built in Python, utilizing standard machine learning libraries including matplotlib (Bisong, 2019; Hunter, 2007), numpy (Oliphant, 2006; Harris et al., 2020), pandas (McKinney et al., 2011), and sklearn (Pedregosa et al., 2011). From sklearn, the RandomForestClassifier, RandomForestRegressor, train_test_split, and accuracy_score played especially important roles, allowing the proper training of the model and display of accuracy for later analysis. In regards to the train_test_split method, the Random Forest implementation in this study used a test size of 0.2 meaning that 20% of the inputted dataset was used as a testing set and 80% was used as a training set. The model was tested using 10, 30, and 100 decision trees to assess the effect of estimator number on the accuracy of the classification, ultimately determining stability of results.

Results

Findings posit that the three most influential factors in accurately deriving collision likelihood were eccentricity, inclination, and radar cross sectional area of the chaser object. In addition, time series handling of data allowed more accurate, organized results. Essentially, time series organization refers to the refitting of the dataset to account for the repetition of event IDs. Information recorded under the same event ID was realigned to fit in the same row, ultimately stretching out the dataset to contain fewer rows.

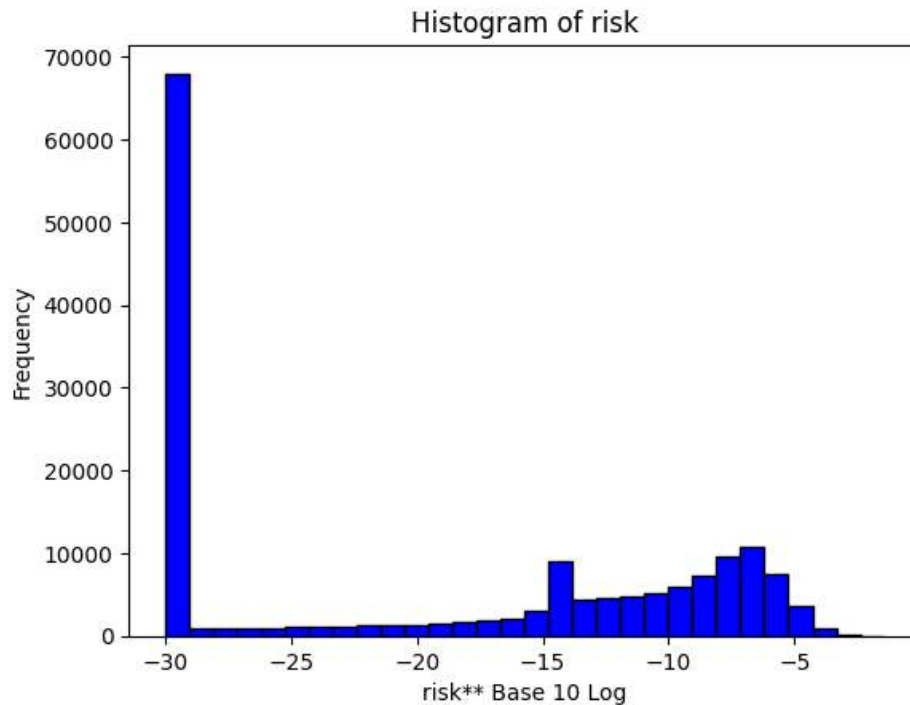


Fig 1: The histogram above depicts the risk on a log base 10 scale. An evident spike in the risk value is displayed around the value of -30 while a gradual increase can be noted from -25 until -7 where it begins to descend.

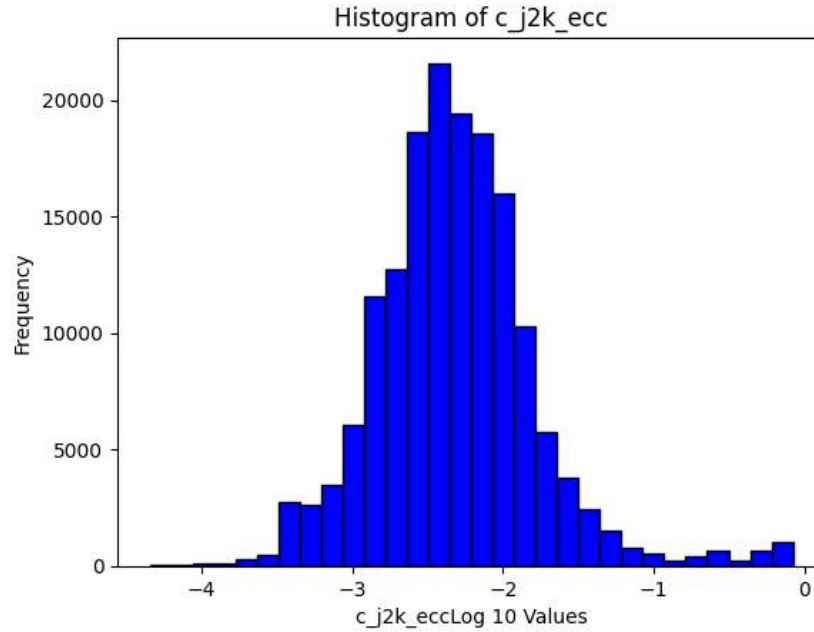


Fig 2: The histogram above demonstrates a peaked distribution of chaser orbital eccentricity, centered around log base 10 of the eccentricity being approximately -2.5. This corresponds to raw eccentricity values of around 0.002-0.005, indicating that most of the orbits of chaser objects are nearly circular. Dramatic declines towards the ends of the figure suggest that there are few highly elliptical orbits.

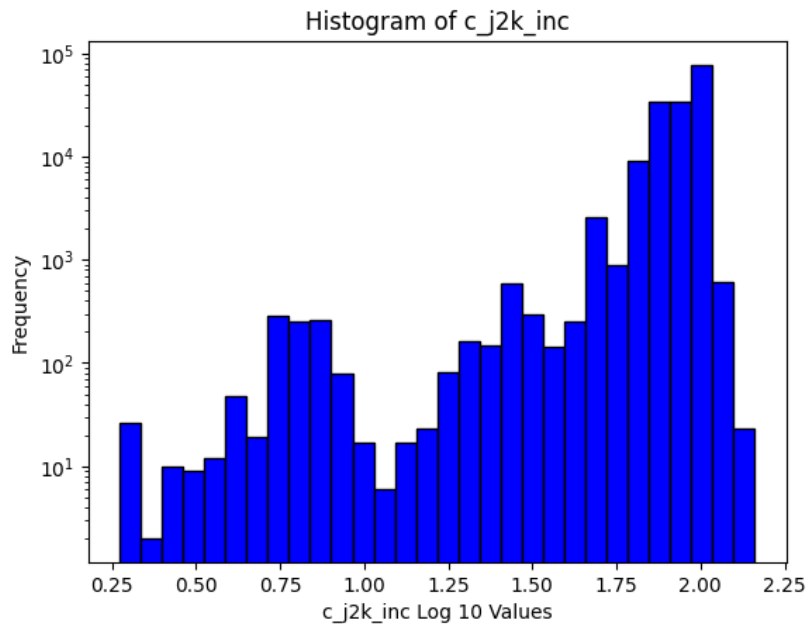


Fig 3: This histogram of chaser orbital inclination displays a strongly right skewed distribution with a maximum peak value around log base 10 of the inclination being approximately 1.9-2. The raw inclination values would thus be around 80 to 100 degrees, suggesting that there is a higher density of sun-synchronous orbits. Lower inclinations appear far less frequently, though the separation between common high and and rare low inclination orbits insinuates that orbital inclination is an extremely discrete variable.

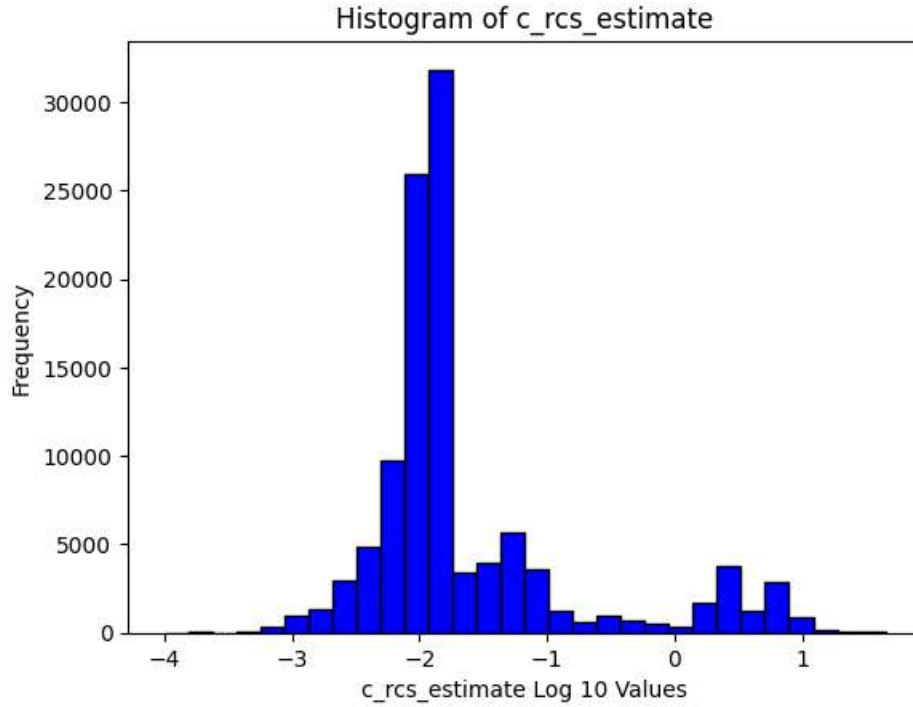


Fig 4: Distribution is sharply concentrated near a log base 10 value of around 2, indicating a raw radar cross-section of approximately 0.01 meters squared. The horizontal expansion of the graph to the right indicates that most chasers are small debris objects while larger objects only constitute a minority of contributions to collision risk.

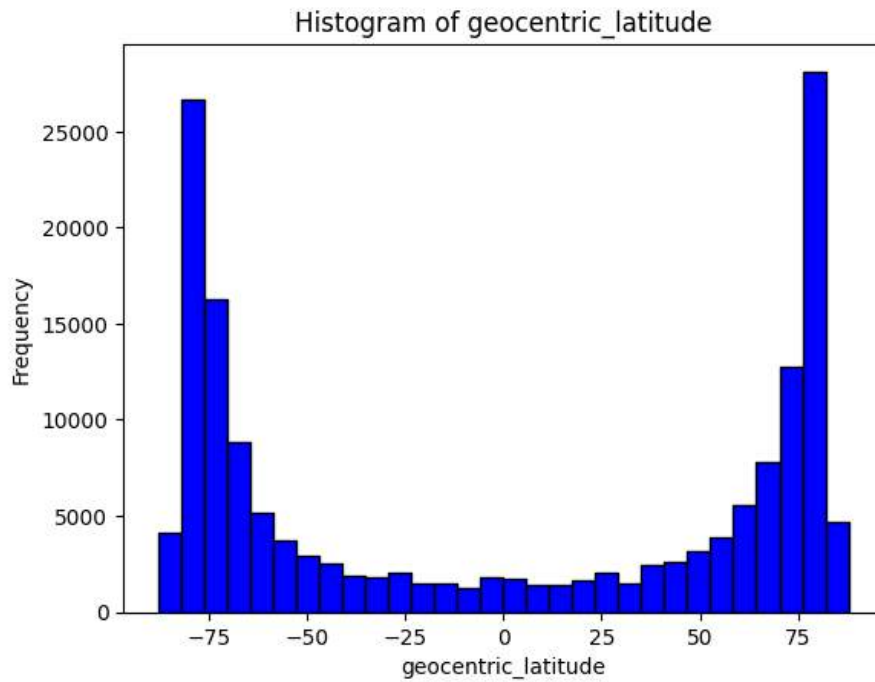


Fig 5: This histogram demonstrates a bimodal structure (peaks in two distinct areas), reaching maximums near ± 75 -80 degrees. This suggests that satellites occupy high-inclination or near polar orbits. A minimum at the equator (0 degrees) indicates that fewer objects exist in a low latitude orbital plane.

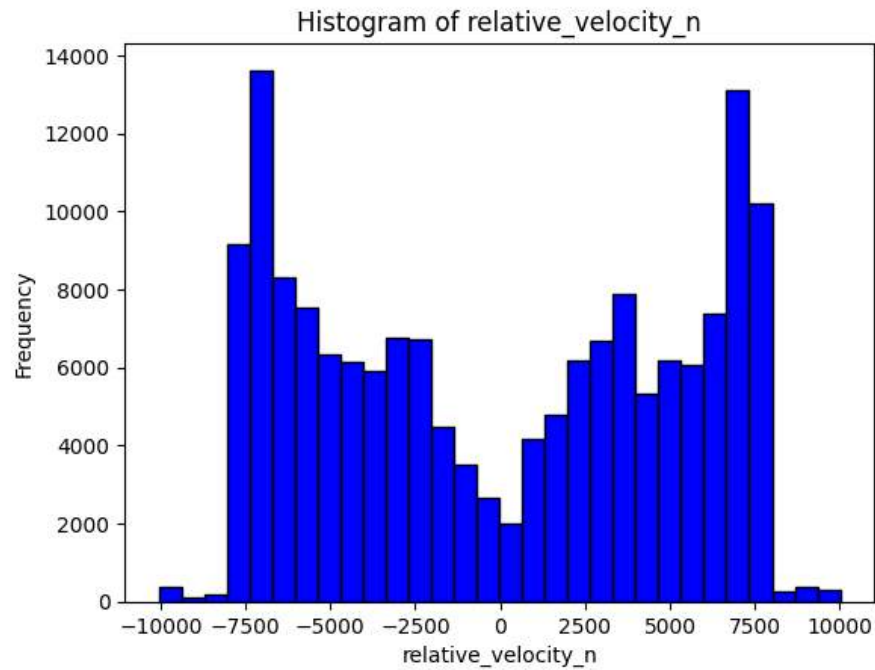


Fig 6: The histogram displays a zero centric distribution with 2 peaks around ± 7000 - 8000 m/s, indicating that there are frequent high speed perpendicular events between the satellite and the considered debris. This allows the conclusion that collisions would have severe impact.

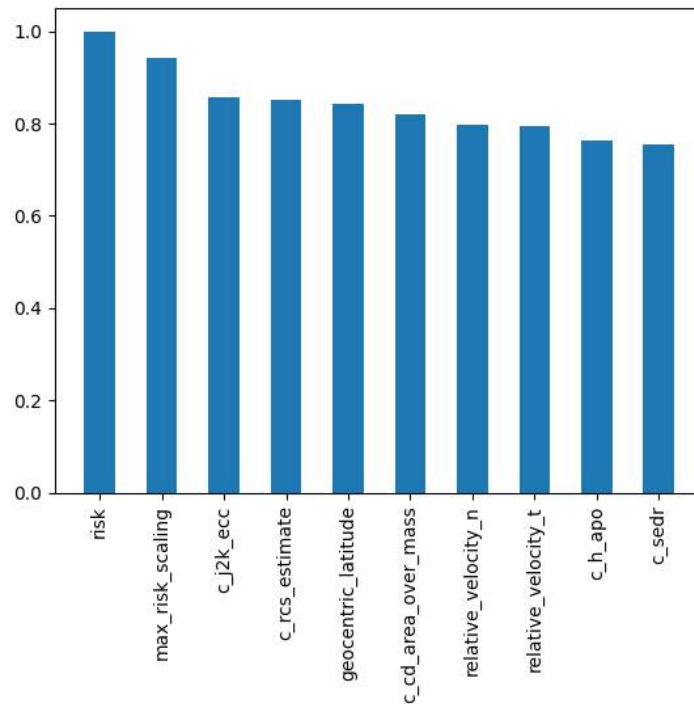


Fig 7: 10 Tree Single Feature Random Forest Feature Importance Chart

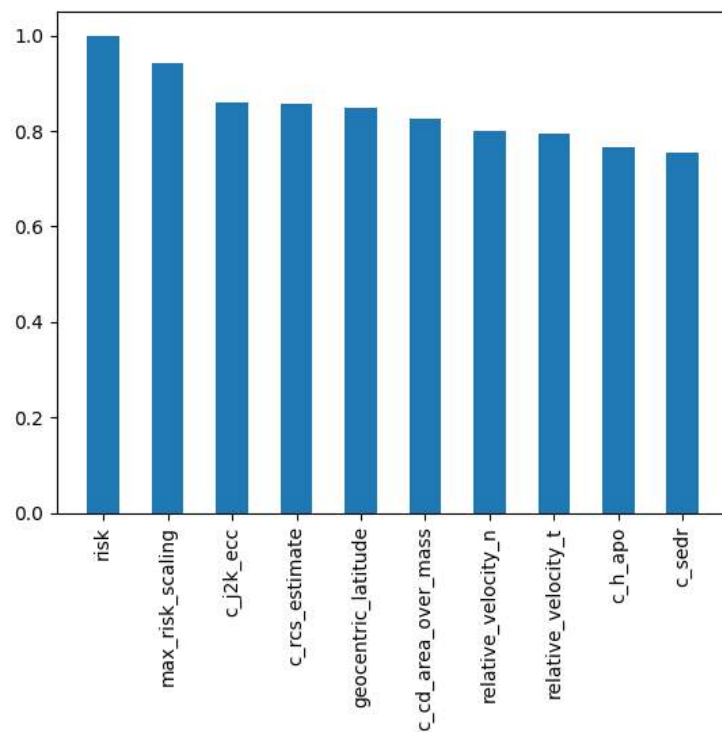


Fig 8: 30 Tree Single Feature Random Forest Feature Importance Chart

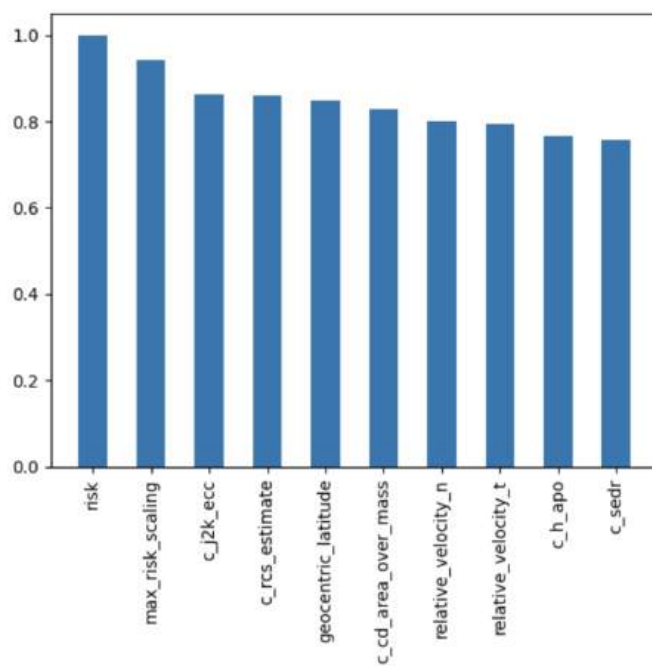


Fig 9: 100 Tree Single Feature Random Forest Feature Importance Chart. Across all numbers of estimators in a SF model, chaser eccentricity and radar cross-sectional area consistently prove to be the most influential variables. The stability of value importance across number of trees indicates the robustness of the single feature prediction model.

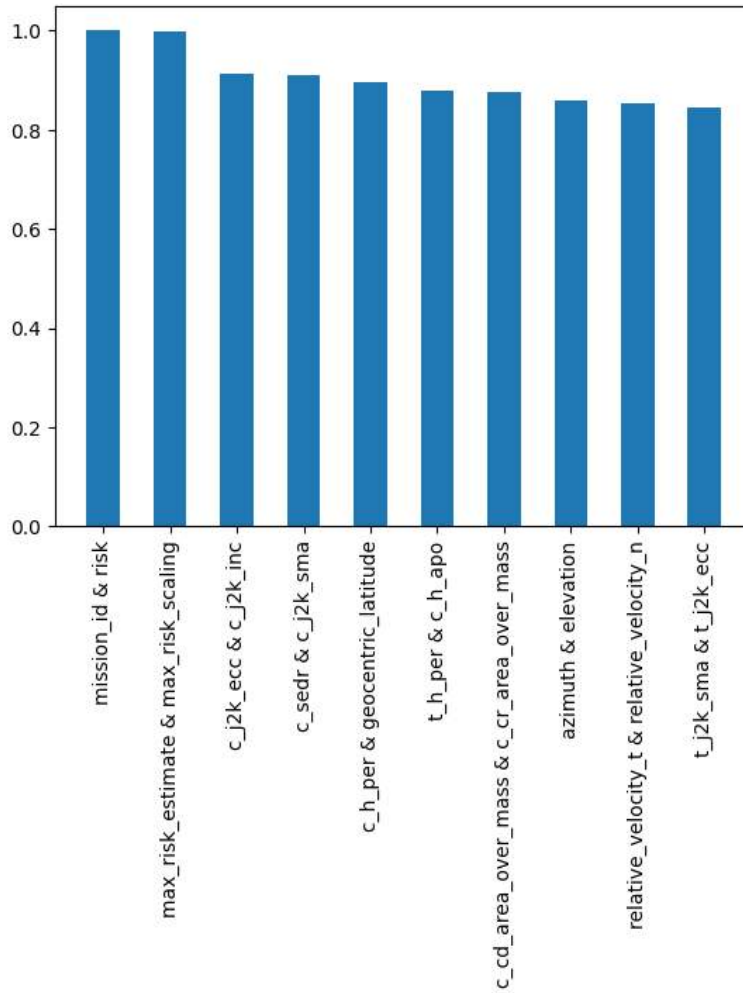


Fig 10: 100 Tree Double Feature Random Forest Feature Importance Chart. Paired features yield eccentricity and inclination to be the most accurate (not considering given risk comparison values from the inputted dataset) while diminishing accuracies are demonstrated for secondary variables (azimuth, elevation, etc.). An increasing estimator count results in some change of the most accurate values changing RCS to inclination.

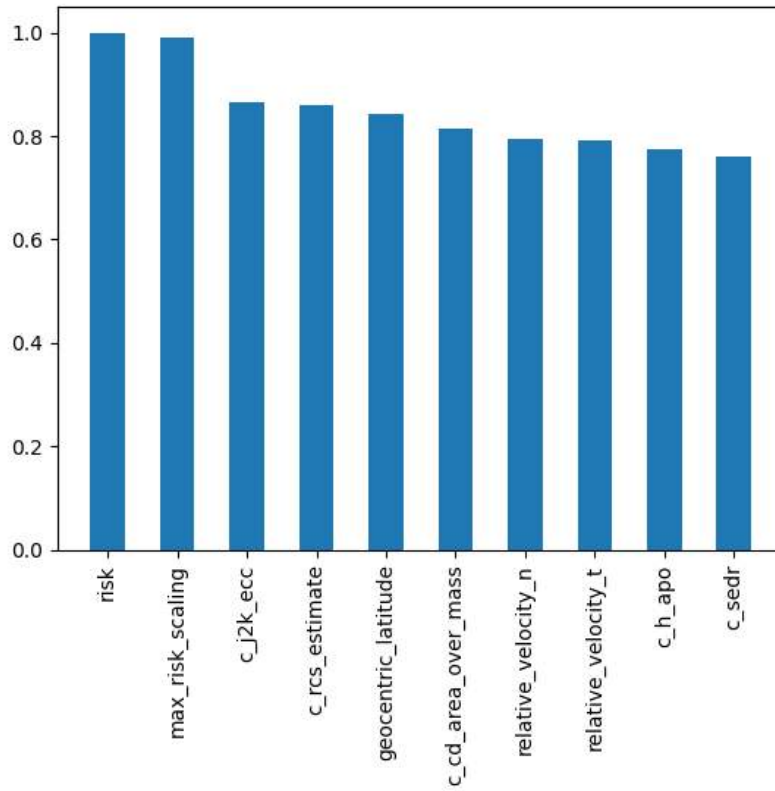


Fig 11: 100 Tree Single Feature Random Forest with Normalized Data Feature Importance Chart. Post normalization, the same variables, eccentricity and rcs estimate remain the most highly ranked accuracies confirming that scale bias did not play a role in previously derived conclusions.

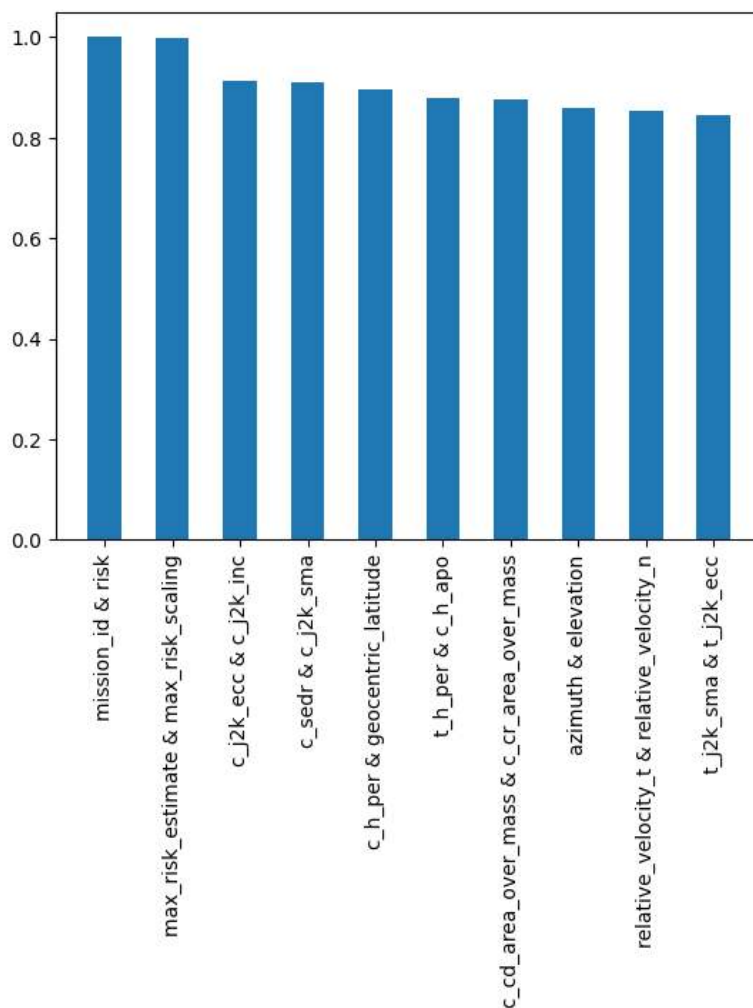


Fig 12: 100 Tree Double Feature Random Forest with Normalized Data Feature Importance Chart. Again normalized parameters do not change the outcome in a double feature analysis, leading to the conclusion that variables revolving around orbital and spacecraft geometry are especially influential when compared to those regarding drag, etc.

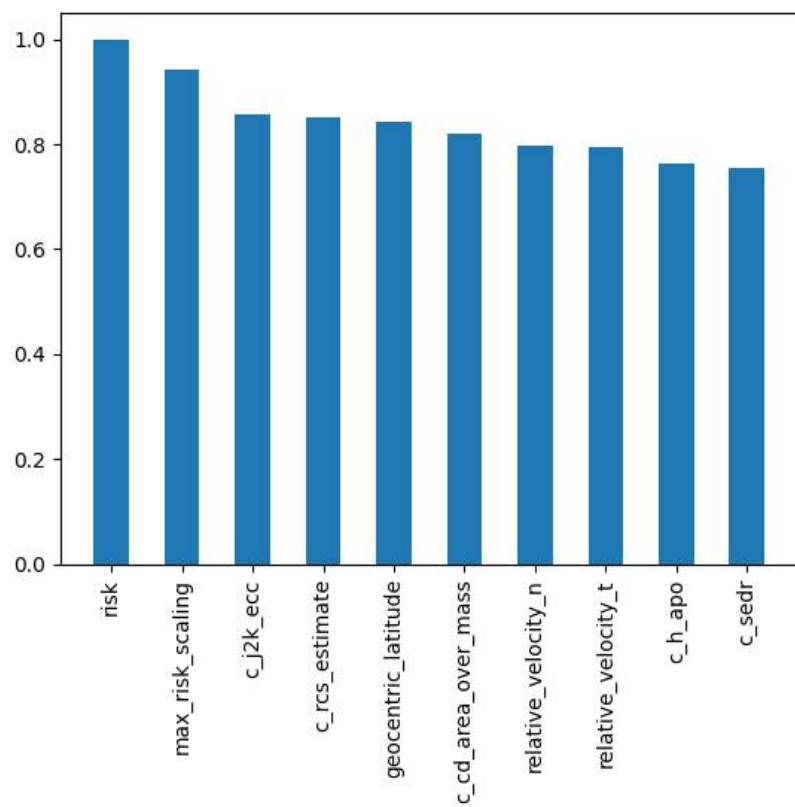


Fig 13: Time Series Adjusted Data Set - 10 Tree Single Feature Random Forest Feature Importance Chart

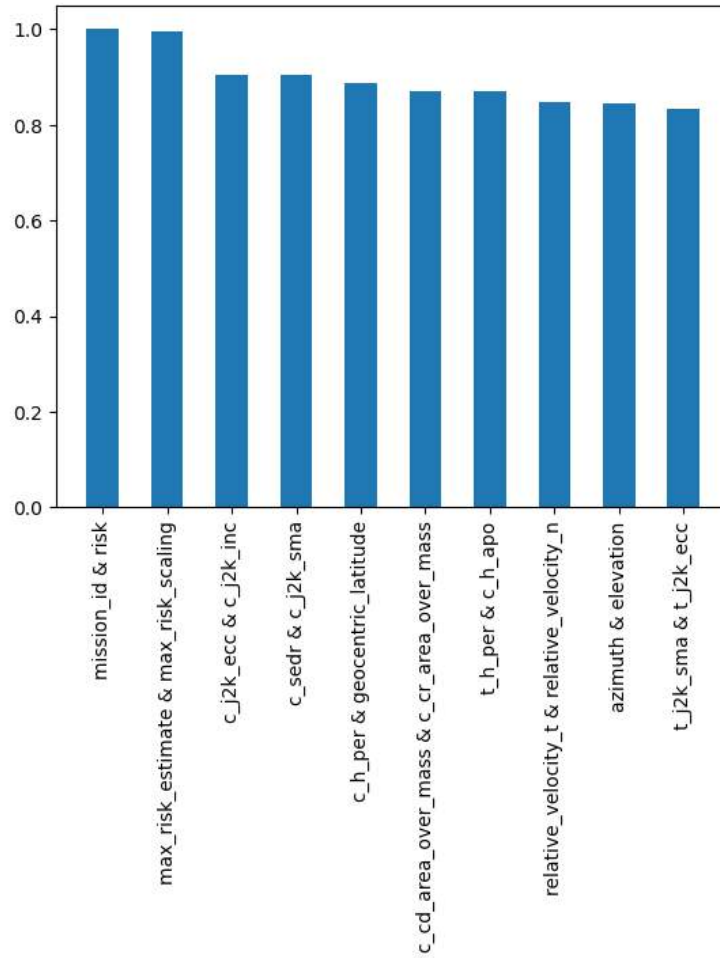


Fig 14: Time Series Adjusted Data Set - 10 Tree Double Feature Random Forest Feature Importance Chart. Time series alignment retains the same results yielding eccentricity and inclination to be the most important in DF analysis and eccentricity and radar cross-sectional area to be the most important in SF analysis though it does sharpen distinctions between accuracy values a noticeable amount.

Discussion

Using Random Forest models, this project found that the most influential factors affecting orbital collision likelihood were `c_j2k_ecc`, `c_j2k_inc`, and `c_rcs_estimate`. To better understand this outcome it is necessary to break these terms down: the preceding letter `c` indicates that the value provided is in reference to the chaser, or the object intended to be avoided by the satellite; the initial `j2k` is a standard inertial reference frame, indicating that the values provided are not in relation to Earth-fixed coordinates but rather to a J2000 spatial setting. A `j2k` frame of reference is Earth-centric, which should be distinguished from Earth-fixed, the former meaning that the origin of the coordinates is the center of the Earth, and the latter indicating that the axes rotate with the Earth, changing as the Earth spins; the third value in these variables refers to a Keplerian orbital element, a set of parameters that describe the shape, size

and orientation of a satellite's orbit. In this case, the three most important factors are the orbital eccentricity of the chaser object, rather its shape, the inclination of the chaser's orbit, and its cross sectional area. Eccentricity (e) describes whether the chaser orbit is circular or elliptical where an $e = 0$ setting describes a circular orbit and $0 < e < 1$ describes an elliptical orbit. Inclination refers to the object's orbital angle measured in degrees or radians, in this case in relation to the J2000 equatorial plane. The radar cross-section estimate (rcs estimate) is a measure of how capable an object is of detection by a radar. In essence, it describes the effective area that reflects back to a radar sensor. A larger RCS indicates an easily detectable object and a smaller one indicates the opposite.

Further expansion upon this project may involve the consideration of gradient boosting, or regression algorithms to compare the robustness of the separate models. In addition, to amend the limitations of a Kaggle computational environment, a more robust model with higher storage and processing capabilities may allow further exploration and comparisons to be made. Beyond the scope of possibility provided by a chosen IDE, limitations to this project include the considered variables, among which atmospheric density was not considered.

Conclusion

Ultimately, this project yielded that eccentricity, inclination, and radar cross sectional area of the chaser object were the three most important factors in predicting the collision likelihood of a satellite in orbit. This conclusion was attained using random forest methods of artificial intelligence implementation within a Kaggle environment and a Kelvins ESA dataset. Through the development of several different random forest models based on normalized and time series altered data, altering the number of trees within these models, and creating feature importance plots, the most significant variables were determined. This research will enable the accurate prediction of satellite collisions to be determined using inputs to influential variables, most specifically and robustly through the eccentricity, inclination, and cross sectional area of the chaser object.

Works Cited

1. European Space Agency. "Space Debris by the Numbers." *ESA Space Debris Office*, European Space Agency, www.esa.int/Safety_Security/Space_Debris/Space_debris_by_the_numbers. Accessed 14 Jan. 2026.
2. Drolshagen, Gerhard. "Impact effects from small size meteoroids and space debris." *Advances in space Research* 41.7 (2008): 1123-1131.
3. Smirnov, N. N., et al. "Space traffic hazards from orbital debris mitigation strategies." *Acta Astronautica* 109 (2015): 144-152.
4. Hiriart, Thomas, et al. "Comparative reliability of GEO, LEO, and MEO satellites." *Proceedings of the international astronautical congress, IAC-09 D. Vol. 1. Paris, France: International Astronautical Federation (IAF), 2009.*
5. Liou, J. C., et al. "Stability of the future LEO environment—an IADC comparison study." *Proceedings of the 6th European Conference on Space Debris. Vol. 723. ESA, 2013.*
6. Vatalaro, Francesco, et al. "Analysis of LEO, MEO, and GEO global mobile satellite systems in the presence of interference and fading." *IEEE Journal on selected areas in communications* 13.2 (2002): 291-300.
7. Vallado, David A., and Paul Crawford. "SGP4 Orbit Determination." *AIAA/AAS Astrodynamics Specialist Conference*, American Institute of Aeronautics and Astronautics, 2006.
8. Vallado, David A., et al. "Revisiting Spacetrack Report #3." *AIAA/AAS Astrodynamics Specialist Conference*, American Institute of Aeronautics and Astronautics, 2006, doi:10.2514/6.2006-6753
9. Rigatti, Steven J. "Random forest." *Journal of insurance medicine* 47.1 (2017): 31-39.
10. Breiman, Leo. "Random Forests." *Machine Learning*, vol. 45, no. 1, 2001, pp. 5–32.
11. Kavlakoglu, Eda. "What Is Random Forest?" IBM, 20 Nov. 2025, www.ibm.com/think/topics/random-forest.
12. "Data." *Kelvins*, kelvins.esa.int/collision-avoidance-challenge/data/. Accessed 26 Jan. 2026.
13. Bisong, Ekaba. "Matplotlib and seaborn." *Building machine learning and deep learning models on google cloud platform: A comprehensive guide for beginners*. Berkeley, CA: Apress, 2019. 151-165.
14. Hunter, John D. "Matplotlib: A 2D graphics environment." *Computing in science & engineering* 9.03 (2007): 90-95.
15. Oliphant, Travis E. *Guide to numpy*. Vol. 1. USA: Trelgol Publishing, 2006.
16. Harris, Charles R., et al. "Array programming with NumPy." *nature* 585.7825 (2020): 357-362.
17. McKinney, Wes. "pandas: a foundational Python library for data analysis and statistics." *Python for high performance and scientific computing* 14.9 (2011): 1-9.
18. Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." *the Journal of machine Learning research* 12 (2011): 2825-2830.

Unmasking Morality: The Role of Moral Judgment in Historical Scholarship

By Luke Cutler

The French Historian Marc Bloch once inquired, “Are we so sure of ourselves and our age as to divide the company of our forefathers into the just and the damned? ...Robespierrists! Anti-Robespierrists! For pity’s sake, simply tell us what Robespierre was” (Gaddis). While quite memorable, Bloch’s sentiment is only a drop in the pond of a long-standing historiographical debate: should historians imbue their works with their personal moral judgments of the past, or should they omit such commentary for the sake of objectivity? Some, such as C.B. McCullagh, contends that historians can maximize impartiality through a stalwart commitment to rational inquiry and the vigilant verification of their conclusions (McCullagh). On the other hand, many assert that, due to the inherent bias of both historical archivists and documents, historical objectivity is a white whale that historians may fanaticize over but will never catch (McCullagh). Despite the allure of both perspectives, the truth is far more complex. Not only are bias and judgment challenging to expunge from historians’ compositions, but, in some respects, moral commentary has the potential to enhance the transparency and reliability of historical narratives. For these key reasons, explicit moral critiques of past institutions and people should be welcomed into historical scholarship.

Admittedly, avoiding judgment and bias in the first place would be a far more attractive alternative. However, this is unfortunately impossible, as humans have been shown to reflexively assess the character and ethics of decisions, events, and people mere milliseconds after first contact. According to a study from Bertram F. Malle of Brown University, humans generally make ‘evaluative judgments’ (determining whether moral stimuli are good or bad) in “300-600 ms [milliseconds] when measuring the earliest electrophysical responses” (Malle). Per evidence cited in the same study, the estimated speed of ‘wrongness judgments’ (determining whether something is morally wrong) is just 800 milliseconds (Malle). Perhaps more shockingly, experiments performed by Princeton psychologists Janine Willis and Alexander Todorov found that “a tenth of a second is sufficient to make a specific trait inference from facial appearance” (Willis). In other words, upon seeing another person for the first time, all it takes is a fraction of a second for us to form conclusions about their character. While Willis and Todorov note that increased exposure time often complicates these conclusions, they affirm how “the judgments are already anchored on the initial inference” (Willis). All in all, these details substantiate how we as humans cannot help making moral or other evaluative judgments of stimuli, given that these reactions are nearly instantaneous and, as the University of Sheffield confirms, “a product of associative...automatic...processes” (Dubljeuk). Regardless of historians’ frequent, vicarious travel to past lives and distant lands, they are just as human as the rest of us and thus are just as predisposed to the snap moral judgments described above.

Having established the inevitability of morally evaluating actions, events, and people, the question still stands: why cannot historians simply exclude their personal biases from their writings? Well, although historians can and do actively strive to remain objective and

even-handed, their judgments likely still manifest in their published work because of the objectivity illusion. The objectivity illusion is a well-known phenomenon in psychology where individuals treat their perception of the world as dispassionate truth and often characterize opposing viewpoints as biased or otherwise invalid (Schwalbe). Considering this phenomenon, no matter how much they strive to avoid moral judgment, historians could inadvertently implant their ethical opinions within their work, reframed as objective fact, thereby misleading both their audience and themselves. This historian's trap--- the failure to eliminate subjective moral commentary while striving for complete objectivity--- is not solely theoretical, being readily observable in prominent historians such as Herbert Butterfield, a historian of English history who extensively contemplated the intersection between ideology and historical interpretation. Like the aforementioned historical objectivists Bloch and McCullagh, Butterfield condemned moral judgment in historical narratives, emphasizing how "Above all it is necessary to resist those who claim for the historian the solemn role of moral arbiter, and particularly those who transfer this ethical preoccupation into the reconstruction of the whole course of ages" (Butterfield). Despite Butterfield's fervent resistance to historians becoming 'moral arbiters,' to parse out the rights and wrongs of the past, he himself was unable to cleanse his own works of his ideological leanings. Specifically, in his 1953 publication 'Christianity, Diplomacy and War,' he claimed that it was "wrong for men to lose faith in the old League of Nations, for though it failed in spectacular manners...many of its activities had the effect of a kind of knitting process, fabricating a tradition of common action" (Andersen). In characterizing these men of the past as wrong to mistrust the League of Nations explicitly because of its role in fostering transnational 'common action,' Butterfield indicates that he believes such political internationalization to be good or beneficial. Therefore, in this excerpt, Butterfield embodies what he purportedly dreads historians becoming: a moral arbiter judging past attitudes based on his personal moral metrics. This in no way intends to cast Herbert Butterfield as a hypocrite but simply serves to evidence how an intense commitment to historical amorality is insufficient to purge bias and moral judgment entirely.

Another dark side of historical objectivity occurs when a desire to remain neutral leads to historical misrepresentation. One historian who has fallen victim to this tendency is E.H. Carr. In keeping with his view of judging historical figures as "the 'principal embarrassment' of the contemporary historian," Carr hesitated to morally assess Stalin and the U.S.S.R, even though "he had ample evidence of Stalin's crimes and yet sought to shroud them in utilitarian calculations about the price of... 'progress'" (Gaddis). By refusing to emphasize or condemn Stalin's atrocities, Carr inherently portrayed him in a more positive light, sugarcoating the cruel brutality that Stalin put on full display with exterminations such as the Holodomor--- a man-made starvation campaign that killed approximately 3.9 million Ukrainians. If, instead of rationalizing them, Carr had elaborated upon or denounced the horrors perpetrated by Stalin's government, those whom Carr influenced would likely have come away with a markedly different assessment of Stalin and the Soviet Union (Applebaum). Thus, E.H. Carr's ambivalence toward Stalin's morality demonstrates how impartiality can breed misrepresentation.

The alternative is not for historians to abandon objectivity altogether, but to explicitly state and include their moral verdicts within their works, alongside the standard fact-grounded narratives of history. Instead of attempting to suppress their judgments, historians should take the time to recognize them and to put these opinions to page (when appropriate) in a format that clearly distinguishes them from more empirical content. If properly incorporated, this structure would not only enable the audience to thoughtfully engage with the author's critiques but it would also arm readers with the information needed to detect areas of the narrative where the author's bias seeps through the cracks.

Overall, not only would adopting moral and ideological judgment as a core feature of historical scholarship not compromise the field, but it may actively increase its integrity and benefit students of history. Whatever its theoretical impact, responsibly integrating moral assessments is certainly a more viable approach for historians than absolutely omitting such biases, which, as this paper has illustrated, is highly unreliable. Morality is a force that pervades both the past and the present, both history and historians. One cannot simply make it disappear. As the historian John Lewis Gaddis puts it, "To purge human behavior of a moral sense is to deny what distinguishes it. You'd be writing histories of schools of fish, flocks of birds, not people" (Gaddis).

Works Cited

- Andersen, Matt. "The Myth of Herbert Butterfield's Political Realism." *Journal of the History of Ideas Blog*. 15 Nov. 2023.
<https://www.jhiblog.org/2023/11/15/the-myth-of-herbert-butterfields-political-realism/>.
- Applebaum, Anne. "Holodomor." *Britannica, Encyclopedia Britannica*, 6 Feb. 2026.
<https://www.britannica.com/event/Holodomor>.
- Butterfield, Herbert. *History And Human Relation*. E-book ed. 1951. Internet Archive.
<https://archive.org/details/HistoryAndHumanRelation/page/104/mode/1up>.
- Dubljeuk, Veljko. "An interesting time for the study of moral judgement and cognition." *University of Sheffield*. 29 Apr. 2015.
<https://icog.sites.sheffield.ac.uk/blog/an-interesting-time-for-the-study-of-moral-judgement-and-cognition>.
- Gaddis, John L. *The Landscape of History: How Historians Map the Past*. Oxford University Press, 2002.
- Malle, Bertram F. "Moral Judgments." *Annual Review of Psychology*. Vol. 72.
<https://research.clps.brown.edu/SocCogSci/Publications/Pubs/Moral-Judgments.pdf>.
- McCullagh, C. Behan. "Bias in Historical Description, Interpretation, and Explanation. *History and Theory*, vol. 39, no. 1. Feb. 2000, pp. 39-66. *Academia.edu*.
https://d1wqtxts1xzle7.cloudfront.net/118525224/HIS4010_McCullagh-libre.pdf?1727686896=&response-content-disposition=attachment%3B+filename%3D-Bias_in_Historica_Description_Interpret.pdf.
- Schwalbe, Michael C., et al. "The objectivity illusion and voter polarization in the 2016 presidential election." *PNAS, Proceedings of the National Academy of Sciences*. 17 Aug. 2020. <https://www.pnas.org/doi/10.1073/pnas.1912301117>.
- Willis, Janine, and Alexander Todorov. "First Impressions: Making Up Your Mind After a 100-Ms Exposure to a Face". *Psychological Science*. Vol. 17, no. 7. 2006. pp. 592-598.
https://cpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/f/3051/files/2021/02/WillisTodorov_PS2006.pdf

Politics and Policy: To What Extent Do Past Health Experiences Affect the Support Levels of a Political Elite in the U.S. Congress By Abeer Shah

Abstract

This paper examines how personal health experiences, specifically contracting Covid-19, influence the policy decisions of U.S. senators. It challenges the generic explanations of legislative behavior such as party affiliation, economic incentives, and public opinion by focusing on psychological and experiential factors. By doing this, it hopes to provide another logical viewpoint. The study asks: To what extent do personal health experiences shape policymaker judgment and voting behavior on health policy? To conduct the research, a highly-advanced classification system was utilized supported by machine learning logic to avoid bias. Public statements, press releases, and voting records are collected and analyzed from before and after senators contracted Covid-19 and each of these sources are categorized on a scale from strong opposition to strong support, allowing for a measurable comparison of changes over time. The study includes four case studies; three senators who contracted Covid-19 and one that did not (a control group). The findings show that getting COVID-19 does not always lead to greater support for health policies. In many cases, there was actually a small to moderate drop in support after infection, especially when the symptoms were mild. The control case, where the senator did not get COVID-19, showed that their policy views stayed mostly consistent over time. These results go against other research that says major health crises usually increase support for health policies. Instead, this study suggests that the relationship between personal experience and political decision-making is more complicated. Overall, the paper shows that personal and psychological factors play an important role in how lawmakers make decisions, and that health experiences can influence policymakers across party lines.

Introduction

In April 2020, as the first wave of Covid-19 brought the world to a standstill, the public received a chilling update: Prime Minister Boris Johnson had been moved to the Intensive Care Unit at St. Thomas' hospital in London. Just days prior, he had been attempting to project an image of business-as-usual, but the virus did not respect political office. As Johnson later described, "it was a tough time that forced him to confront the reality of the pandemic from a victim's view" (Express UK News). That moment, when a head of government became just another patient battling the deadly disease, the illusion was shattered that leadership positions offered protection. Instead, it highlighted the absolute vulnerability of any human being to Covid-19. That moment was a stark reminder to the world that politics, health, and various outside factors are inadvertently intertwined.

Relatedly, this paper asks: What affects the way politicians decide to vote, specifically in the United States. Whether the deciding factors are monetary gain, personal danger, or popularity growth, the impact is seen even when not explicitly stated. These factors affect how senators vote because it changes their perception of a topic and their personal benefit. This research will focus on how past medical experiences affect policymaker decisions, and more specifically, Covid-19. Covid-19 was a global pandemic that caused social and economic distress worldwide. The

research question that the article is focusing on is, “To what extent do psychological factors and past health experiences shape legislator decisions and judgement regarding health policy?” This paper argues that personal health factors significantly shape policymaker voting behavior instead of focusing on factors such as economic reasons because health is a cross-ideological factor (Grant et al.). While money, economic reasons, and other factors usually connect to party alignment and ideological blocs, health factors can end up cutting across these party lines. For example, during the Covid-19 pandemic, policymakers from different parties shifted voting patterns due to personal risk level and family health concerns and the lines between parties became blurred. This gives a more human explanation of voting behavior and a factor that affects all policymakers, liberals and conservatives alike.

In the following sections of my paper, I will first cover the literature review, then the methodology, and after that my four case studies on various senators. Furthermore, after my case studies, I will provide an analysis of my results, my discussion of how my paper fits into future research, and the limitations of my paper.

Literature Review

This research sits at the intersection of politics and medicine. Authors such as Dickson and Yildirim focus on how critical issues and global shocks correlate with medicine and how they affect policymaker decisions. The article about the effect of Covid-19 on policy viewpoints investigates with a difference-in-differences design on how political elites responded to global shocks. The researchers did this by using a before and after comparison of elite opinion, seeing how their opinions were before a global shock and after. This article found that Covid-19 caused approximately a 30% reduction in legislators’ expressions of opposition to health policies. While this article focuses on medicine and more especially the effect of global shocks on policymaker decisions, the next article did research more generally into the demographic information of legislators. My article builds upon Dickson and Yildirim’s article since it focuses on collecting a more specific result rather than the other article’s generalized conclusion (Dickson & Yildirim). Additionally, Emrich et al. found that voters born in close proximity to one another tend to exhibit similar voting behaviors during roll call voting, even when accounting for different partisanships. The study was conducted by combining birthplace location data and roll call voting data and analyzing them side by side, seeing if birthplaces in close proximity had any visible effect on voting behaviors. Additionally, they also explored how a lawmaker’s birth county affected their support on agricultural policy. To do this, they measured how agricultural a lawmaker’s birth county was (using USDA Census of Agriculture data on cropland acres near the member’s birth year) and linked that to the lawmaker’s later voting behavior on agricultural policy. Then, using statistical models (including regression and an ideal-point analysis of agriculture-related roll-call votes), they tested whether being born in a more agricultural area predicted stronger support for farm protection policies. They found that the agricultural composition of members’ birthplace influences their support for agricultural policy and protection. In conclusion, their findings suggest that the personal history of lawmakers has a direct impact on how they represent their constituents. This conclusion is directly connected to my project since it provides a basis of how anything in a lawmaker’s personal history can

influence their vote, and my focus is on how health experience in a lawmaker's history can influence their vote. The last study explores how demographics affect how U.S. state legislators vote in the world of medicine.

Pilar et al. focuses on how mental health parity legislation support is different for legislators with various demographics. Utilizing a previously conducted cross-sectional survey of 475 U.S. legislators, the researchers conducted various statistical tests and analysis to find their results. They found that support for parity laws was, "generally higher among females, more liberal legislators, legislators in the Northeast region of the country, and those who had previously sought treatment for mental illness" (Pilar et al.). These findings provide a critical background and structure to the researcher's study, since it helps understand how demographics and outside factors influence voting and support in the broader medical world (Pilar et al.).

Methodology

To collect their results, this paper will use a machine learning model (ChatGPT -5.2) First, 10 sources - social media posts, press releases, interviews, etc. that show a senator's opposition or support of a health mandate before or after contracting Covid-19. Each item will be coded into one of five categories: weak opposition, strong opposition, strong support, weak support, and neutrality. This is where ChatGPT is used to sort all the sources in order to prevent human bias. Each category holds its own percentage value - strong support is 100%, weak support is 75%, neutrality is 50%, weak opposition is 25%, and strong opposition is 0%. Lastly, each senator's 10 pieces of evidence will be split in half - five will be from before the senator contracted the disease, and five will be from after. The researcher will use at least one source to determine when the senator contracted the disease and include them in their final results. After all 10 items are classified for each senator, then, the researcher will find the average percent of the five sources from before the disease (Covid-19) was contracted and the five from after it was contracted. These results show important data of how a senator contracting Covid-19 affects the way they vote, support mandates, pieces of legislation, and bills relating to Covid. When presenting these results, the researcher will provide the severity of the senator's Covid-19 (when ranked on a scale from 1 - 5 by ChatGPT), and when they were infected to provide an external factor. Furthermore, for the control group (senator Patty Murray), the study was conducted differently. To find her percent change in support over time, the researcher collected sources in increments of around two months, analyzing evidence from the years 2020 - 2022. This allowed the researcher to identify how her support changed over time on a controlled scale. Tim Kaine, Bill Cassidy, and Susan Collins were chosen because they represent different backgrounds and perspectives. This is shown by including both men and women, as well as members of different political parties. Additionally, Patty Murray was included as a control to show how a lawmaker's opinion on a disease might change over time, even if they do not have personal experience with it. By selecting people from different parts of the United States, the analysis accounts for how COVID-19 affected different areas, which can influence political positions. Their different professional backgrounds—such as Bill Cassidy's medical experience—provide more insight into how specialized knowledge helps shape responses to a health crisis. Including both moderate and more partisan figures also allows for a better

comparison of how political beliefs impact support. Overall, this selection makes the analysis stronger by including various perspectives while keeping the project focused and manageable.

Case Studies

Case Study 1: Tim Kaine (got covid-19)

The example of Senator Tim Kaine illustrates how a senator's experience with getting sick with Covid-19 influenced them to vote more positively on Covid-19 policy. Senator Tim Kaine first contracted Covid-19 around May, 2020; it was unique since he contracted long covid, which induced persistent, neurological symptoms for many years. This information was collected through analyzing various sources - for example, Senator Kaine stated in a press release in May 2020, "My wife Anne and I recently tested positive for antibodies to coronavirus" (ABC News). Even though Senator Kaine tested positive, his symptoms were mild and showed no severe short-term effects. After he recovered, it was observed that Senator Kaine started supporting Covid-19 policies a little less than before. This result was determined by categorizing ten total sources from before and after Kaine received Covid-19, and it was found that he supported policies 95% and after, he supported policies at an average of 90%. One possible explanation for this result is that Senator Kaine experienced mild Covid-19 and believed it to be less of an issue. Though, it is also possible that this result was a product of random circumstances and that the pieces of evidence didn't provide a comprehensive picture of Senator Kaine's opinions. In conclusion, Senator Kaine was found to support Covid-19 mandates a little less after contracting mild Covid-19, though this conclusion may be incorrect due to pieces of evidence not painting the full picture of Kaine's opinions.

Case Study 2: Bill Cassidy (got covid-19) - an example of a republican who supported mask mandates

Bill Cassidy's example shows a Republican who supported mask mandates and how their opinions changed, making them less supportive after contracting the disease. Cassidy first announced he tested positive for Covid-19 on August 20th, 2020. His spokesman stated that Cassidy was suffering from "mild symptoms that started this morning" and he quarantined at home (PBS News 2020). In fact, Cassidy continued to work from home which showed that his Covid-19 was nowhere near extreme. Once Cassidy recovered, it was observed that Senator Cassidy decreased his support for Covid-19 policies, barely. This conclusion was made after examining and categorizing 10 sources about his support for these policies, five from before he contracted Covid-19 and five from after. More specifically, results showed that his support went down from 95% to 85% after he recovered. The sources used to come to this conclusion included official press releases, news articles, and government voting records. Same as Senator Kaine, while Cassidy may have decreased his support of Covid-19 due to experiencing a mild version, a more likely conclusion is that the sources provided a limited amount of information about his real viewpoints.

Case Study 3: Susan Collins (got covid-19)

Case Study number 3, of Susan Collins, shows a contrasting example to Bill Cassidy of a Republican that ended up being more opposed to Covid-19 policy after contracting Covid-19, unlike him. She first contracted Covid-19 around April, 2022. Her office stated that “[she] is currently experiencing mild symptoms. The Senator will isolate and work remotely in accordance with CDC guidelines” (Collins Senator Organization). While this information may not be the most accurate, it is closest we can get to finding out the severity level of a Senators’ disease. Because, unlike Bill Cassidy who worked from home or Tim Kaine who continued to publicize his doctor visits, there was no real evidence that showed Collins really had mild symptoms. After completing their study, the researcher discovered that Susan Collin’s support level went down from 85% to 50%. Before she contracted Covid-19, she strongly supported many policies, and only showed mild resistance to few Covid-19 policies. As seen in November 2020, Collins joined a bipartisan group that were pushing Covid-19 stimulus deals [relief and money given to families during the pandemic] (Treene). One of the few mandates she opposed before contracting the disease was an Executive Order from President Biden forcing private employers to vaccinate their employees. (Senator Collins’ Website Editors). Through this example, I was able to come to the conclusion that before contracting Covid-19, she showed heavy support but not one hundred percent support for policy. After contracting Covid-19, it was observed that her support levels drastically decreased. Instead of supporting most Covid-19 mandates, she ended up opposing many of them and staying neutral on others. Rarely was it seen that Senator Collins, after contracting the disease, voted in favor of a Covid-19 policy. For example, she began opposing vaccine mandates to government organizations, changing from before when she only voted against vaccine mandates to private organizations (Susan Collins’ Website). The main explanation for this change is that when experiencing mild Covid-19 symptoms, she believed that it wasn’t too impactful - so she stopped voting for health policies correlating to the disease.

Case Study 4: Patty Murray (no covid-19)

Patty Murray, as a control example of a Senator who was never infected with Covid-19, shows us that viewpoints remain constant. She first began supporting Covid-19 policy strongly around April 2020, as seen in a \$30 billion plan she led for constant testing and contact tracing (Fernandez). Emerging as one of the strongest voices behind federal health mandates and policies, she demonstrated major support throughout the early pandemic response period (April 2020 - January 2021.) By 2021 however, her support had reduced a bit, and she began moving away from Covid-19 policy and focused on other laws. Though she still argued for many policies, the amount she supported was not nearly as much as 2020. In 2022, seen in the three sources analyzed, her support levels went back up to the degree of 2020. As seen in a Senate Press Release, Patty Murray was urging Congress to support Covid-19 bills aimed at reconstruction. Overall, throughout all three stages of the pandemic, her support levels remained extremely high. This control shows us that a Senator without a personal connection to a problem normally maintains their viewpoints.

Discussion

Based on the findings, the researcher recommends that doctors and public health advocates prioritize engaging senators before they are personally affected by a disease or infection, since this is when their policy positions are most likely to be shaped by raw evidence rather than personal experience. As seen in the research, personal experience is very likely to reduce their support for health policy regarding that infection. This early outreach allows medical professionals to present data, patient case studies, and public health impacts in a more neutral environment, increasing the likelihood that legislators will support the health policies. Once a senator has been directly impacted by a disease however, their perspective may shift into a more emotional or individualized direction, which can narrow their viewpoint to solutions that only seem practical for their specific infection.

Before the study, the researcher's hypothesis recommendation was that doctors should engage with senators who had already been infected with a disease to ensure they understood the severity. By understanding the severity, they would be able to make more educated and accurate decisions. However, as the results show, this hypothesis was proven wrong and it is more effective for doctors and policy advocates to engage with senators before they are infected with a disease. Referring back to the literature review, the main paper this research is correlated to, Dickson & Yildirim, is contradicted by this paper. In Dickson & Yildirim's research, they found that senators are more likely to support health policy after contracting a disease - a stark contrast to the researcher's conclusion. An educated guess about the cause of this difference in results is the group analyzed, Dickson & Yildirim focused on a broader group in the senate, while this paper focused on senators more specifically related to health policy.

Limitations

While the research was constructed in a way to provide strong, correct conclusions, some limitations still arose. First, a primary limitation of this study was the relatively small sample size of only four senators. Although this may appear to not fully represent the broader population of senators, these individuals were chosen intentionally since they reflect a common demographic within the health committee. To increase balance and reduce bias, I included two women and two men, with representation from both major political parties. This structure allowed for a basic level of diversity in demographics and perspectives, even if it didn't capture as a whole all 23 senators' views that make up the entire committee.

Another limitation arose from the scope of the analysis. More specifically, the research was unable to incorporate certain factors that potentially could have been influential, such as the number of times a senator was infected with Covid-19. These factors could have a reasonable influence on personal perspectives and policy support. However, due to constraints of available and accurate data, these variables were not fully encompassed into the analysis. Despite this, the overall impact of this limitation on my conclusions is minor, since my results focus on the percent change in support rather than establishing a direct relationship between personal health experiences and policy decisions. Furthermore, this study does not account for other external influences that may shape a senator's stance, such as political pressure, constituent opinions, media coverage, or evolving public health guidance over time. These elements are difficult to quantify but could play a significant role in shaping outcomes.

Though, this study acknowledges them setting up future studies to focus on these factors as well.

By clearly acknowledging these limitations, the analysis remains transparent and grounded. At the same time, these limitations emphasize opportunities for future research since they are acknowledged and discussed in this study. Expanding the sample size, incorporating additional demographic and influential variables, and analyzing a broader range of external influences would have allowed for a more comprehensive understanding of this study's results.

Works Cited

- Collins, Susan. "Collins, Colleagues File Challenge to Vaccine Mandate for Businesses." Senator Susan Collins Website. November 18, 2021.
<https://www.collins.senate.gov/newsroom/collins-colleagues-file-challenge-vaccine-mn-date-businesses>
- Collins, Susan. "Senator Collins Tests Positive for COVID-19." Senator Susan Collins Website. April 7, 2022.
<https://www.collins.senate.gov/newsroom/senator-collins-tests-positive-for-covid-19#:~:text=Washington%2C%20D.C.%E2%80%94The%20office%20of,%23%23%23>
- Dickson, Zachary P., and Tevfik Murat Yildirim. "The effects of COVID-19 infection on opposition to COVID-19 policies: Evidence from the US congress." *Political Communication* 42.1 (2025): 127-150.
- Emrich, Colin, Hillary Style, and Ryan J. Vander Wielen. "From cradle to congress: the effect of birthplace on legislative decision-making." *Political Science Research and Methods* (2025): 1-26.
- Grant, M.D., Markowitz, D.M., Sherman, D.K. et al. Ideological diversity of media consumption predicts COVID-19 vaccination. *Sci Rep* 14, 28948 (2024).
<https://doi.org/10.1038/s41598-024-77408-4>
- MacRae, Grace. "Boris Johnson's treatment while in intensive care is revealed as condition deteriorates." *Express News UK*.
<https://www.express.co.uk/news/uk/1265855/Boris-Johnson-intensive-care-coronavirus-update-Prime-Minister-treatment-COVID-19-latest>
- Pilar, Meagan, et al. "An examination of factors affecting state legislators' support for parity laws for different mental illnesses." *Community mental health journal* 59.1 (2023): 122-131.

Cash for Steps: Assessing the Impact of Financial Incentives on Physical Activity in Adolescents

By Yonatan Simcha Rosenberg

Abstract

Financial incentive-based fitness applications have become popular tools to promote physical activity, but questions remain about their long-term effectiveness and optimal design, especially with adolescents, a population at high risk for inactivity and underrepresented in previous research. This study investigates whether financial incentives increase physical activity in adolescents, whether effects continue after incentives are removed, and how incentive magnitude and framing influence outcomes. A longitudinal study was conducted with 26 adolescents aged 14-19. Participants tracked weekly step counts across three phases: a two-week baseline, a two-week incentivized intervention, and a two-week post-intervention period. Participants were randomly assigned to one of four conditions: no incentive (control), small gain-framed incentive, large gain-framed incentive, or small loss-framed incentive. Changes in step counts within and between groups were analyzed, while post-study survey responses provided additional context. During the intervention phase, incentivized groups increased step counts, while the control group experienced declines, suggesting that even modest financial incentives can increase adolescent physical activity in the short term. More significantly, while activity declined across all groups after incentives were removed, previously incentivized participants exhibited smaller declines than the control group, providing tentative evidence of behavioral persistence. Larger incentives produced slightly stronger effects than smaller incentives, while loss framing showed no consistent advantage. However, measurement limitations, ceiling effects, and cost considerations highlight the need for better incentive design. Additionally, the small and unrepresentative sample underscores the need for further research in more diverse populations.

Introduction

In recent years, a growing number of mobile applications have emerged that aim to reshape personal fitness by offering financial incentives to users who meet specific activity goals. One of the most widely known examples is Sweatcoin, which tracks users' steps and converts them into a digital currency that can be redeemed for goods or services in its marketplace and which, according to the app's official website, had over 200 million users as of January 2026 (Sweatcoin). StepBet, another popular platform, takes a slightly different approach by letting users bet on themselves, with participants paying to enter "challenges" and winning cash if they manage to reach personalized targets over the course of a given period (StepBet).

Though these apps are marketed primarily as tools for individual self-improvement, they may also carry significant public health implications. The World Health Organization reports that nearly one third of the global adult population currently fails to meet recommended levels of physical activity, a figure projected to rise to 35% by 2030 (Nearly 1.8 Billion). These findings are concerning given that insufficient physical activity substantially increases the risk of

non-communicable diseases, including breast and colon cancers, diabetes, and cardiovascular disease, as well as mental health conditions such as depression (Santos et al.). Beyond this massive human toll, widespread inactivity imposes a major financial burden on public health systems. A 2022 study using a population attributable fraction approach estimated that without an increase in global physical activity levels, nearly half a billion new cases of preventable major noncommunicable diseases will happen worldwide by 2030, generating direct healthcare costs of approximately \$47.6 billion per year (Santos et al.). And it should be noted that that figure doesn't include indirect costs like loss of productivity due to illness (Santos et al.). In light of these findings, financial incentives present a potentially compelling public health tool. If they are effective in increasing physical activity, they could both reduce preventable disease and associated health-care costs. This raises a critical question: do financial incentives like the ones offered in popular applications actually succeed in motivating individuals to become more physically active?

At first glance, the answer may appear to be yes. In a longitudinal study, Elliott et al. (2019) examined changes in physical activity in users of the app Sweatcoin. Analyzing stepcount data from nearly 6,000 users over a 6 month period, the authors reported an average 19% increase in daily steps, with particularly large effects being observed among previously sedentary and overweight users. They concluded that continuous app-based financial incentives may be effective in motivating increased physical activity, even with populations that were traditionally less responsive to exercise interventions (Elliott et al.). An even larger effect was observed by de Buissonjé et al. (2023), who conducted a large-scale longitudinal evaluation of the app StepBet. Drawing on data from over 72,000 participants, the study found an average increase of 31% in daily steps during six-week challenges, with participants who successfully completed challenges increasing their activity by 44% (de Buissonjé et al.). These findings are fairly consistent with a broader body of research on financial incentives for physical activity, even beyond digital platforms. Strohacker et al. (2014) conducted a systematic review of randomized controlled trials examining the effects of financial incentives on physical activity (Strohacker et al., *The Impact of Incentives*). They found that although unconditional incentives, such as free gym memberships, showed no significant effects on physical activity, conditional incentives (i.e. incentives contingent on meeting specific activity targets) demonstrated robust positive effects, particularly when tied directly to physical-activity behavior rather than mere attendance (Strohacker et al., *The Impact of Incentives*).

However, some important caveats remain. One major question is whether financial incentives can produce lasting behavioral change, or if they merely trigger a temporary spike in activity. The habit formation hypothesis suggests the former, arguing that by providing initial motivation, external incentives encourage people to behave healthily. Over time, people get used to these behaviors and make them part of their daily routines, resulting in the individual continuing the healthy behavior even after the incentives are removed. Meanwhile the motivation crowding hypothesis posits that external incentives can crowd out intrinsic motivation by shifting the perceived reason for acting from personal satisfaction to financial gain. When individuals get

used to doing the behavior primarily because of the incentive rather than due to its inherent value, their interest in continuing without an incentive may decrease. By that logic, financial incentives may inadvertently signal that a behavior is only worth doing for a price, weakening the internal drive necessary for sustained engagement. There is ample empirical evidence to support both views. Examining the long-term effects of financial incentives on health screenings, Jones et al. (2024) found that initial financial incentives not only increased immediate participation but also significantly raised the likelihood of future engagement (Jones et al.). Specifically, individuals induced by incentives to complete a screening in the first year were 36% more likely to complete a screening in the second year and 33% more likely in the third year, consistent with the habit formation hypothesis (Jones et al.). Yet Carrera et al. (2019) found that modest financial or material incentives for new gym members produced only small short-term increases in attendance and had no effect on post-incentive visits, indicating that the incentives did not lead to lasting behavior change, in line with the motivation crowding hypothesis.

Another key question is whether the amount of the financial incentive has any effect on the behavioral change. Intuitively, bigger incentives might be expected to produce greater responses, and much of the evidence supports this expectation, at least in the short term. In a meta-analysis of 29 studies exploring the impact of financial incentives for physical activity, Salmani et al. (2025) found that incentive size had a clear influence on the results, with larger incentives being associated with greater short-term increases in behaviors such as daily step counts. However, evidence suggests that even modest incentives can produce a considerable level of behavioral change and, given their lower cost, they might offer a more sustainable approach to public health. For example, in a 10 week longitudinal study, Strohacker et al. (2015) found that inactive university students who were offered incentives of just \$5 per week demonstrated significant increases in exercise-related energy expenditure (Impact of Small Monetary Incentives). Moreover, larger financial incentives may not produce greater long-term change. In fact, Moller et al. (2014) found that participants who received lower financial incentives experienced increases in how much they enjoyed doing the healthy behavior, while this effect was reversed in those who received higher incentives. Beyond the amount of the incentive, the way it is presented could also influence outcomes. The concept of loss aversion posits that individuals experience losses more acutely than equivalent gains, meaning that they are often more motivated to avoid losing something they already have than to acquire something of comparable value. Consistent with this theory, Patel et al. (2016) found that overweight and obese adults who were given money up front and then lost portions of it for failing to meet step goals achieved their targets more frequently than those who were given gain frame incentives and promised money after completing the step targets. However, further research is needed to determine how generalizable these findings are among different populations.

To address these unanswered questions, the present study examines how financial incentives influence physical activity among adolescents, a population that has been underrepresented in prior research despite being among the most at risk for inactivity (World Health Organization. New WHO-led Study). Using a six-week longitudinal design, this study

tracked participants' step counts across three phases: a baseline period without incentives, an intervention period with incentive-based conditions, and a post-intervention period to assess behavioral persistence once incentives were removed. Four experimental groups were compared: a control group with no incentive, a small gain-framed incentive, a larger gain-framed incentive, and a small loss-framed incentive. Participants were also asked to complete a short followup survey to assess the accuracy of measured steps as a proxy for fitness and gain qualitative insights into the factors shaping their fitness decisions.

The study is limited by its small, convenience-based sample, which may reduce the generalizability of the findings. Nevertheless, it seeks to offer fresh insights into how financial incentives can be strategically applied to promote physical activity among adolescents, a high-risk and underrepresented population, with potentially valuable implications for public health.

Methodology

Research Design

This study used a six-week longitudinal experimental design to examine the effects of financial incentives on physical activity among adolescents. The study consisted of three phases: a two-week baseline period without incentives, a two-week intervention period in which some groups were given incentives, and a two-week post-intervention period during which incentives were removed. The design enabled comparisons both between groups with different incentive conditions and comparisons within the same group over time, including before, during, and after the incentivization.

Participants

The sample consisted of 26 adolescents between the ages of 14 and 19, recruited through convenience sampling. Participants were primarily students attending private schools. While this sample is not fully representative of the general adolescent population, particularly with respect to socioeconomic background, it nevertheless provides insight into how teenagers, a group that has been underrepresented in prior research, respond to financial incentives within the context of fitness.

Procedure

Participants tracked their daily physical activity using a step-tracker application on their smartphones throughout the six-week period of the study, submitting weekly step totals at the end of each week. During the first two weeks, all participants received identical instructions and no incentives in order to establish baseline activity levels. During weeks three and four, a random group generator tool was used to assign participants to one of four experimental conditions. All participants were given a target of 70,000 steps per week. One group received a message encouraging them to meet the step target, but without any incentives. A second group was

offered a \$5 weekly incentive for meeting the target (gain-framed). A third group was offered a \$50 weekly incentive for meeting the target (gain-framed). A fourth group was informed that they had been provisionally awarded \$5, which would be forfeited if they failed to meet the target (loss-framed). Incentives were contingent solely on whether participants exceeded the weekly step target. For the final two weeks of the study, participants were informed that there would be no incentives or special targets. This post-intervention phase was included to observe changes in behavior following incentive removal. At the conclusion of the study, participants completed a voluntary anonymous survey.

Measures and Data Analysis

The primary independent variable was incentive condition during the intervention phase. The primary dependent variable was the total number of steps recorded per week by the step-tracking application, which was used as a proxy measure for overall physical activity. Weekly step totals were compared across baseline, intervention, and post-intervention phases to evaluate changes in physical activity associated with the presence of incentives, as well as their magnitude and how they were framed. Survey responses provided secondary measures used to contextualize the primary outcome. Participants rated the extent to which their recorded step totals accurately reflected their actual steps walked and overall physical activity on a 1-5 Likert scale and explained their ratings in optional free responses. Additional survey items assessed participants' perceptions of financial incentives as a motivator for physical activity.

Ethical Considerations

Informed consent was obtained from all participants prior to participation. Survey responses were collected anonymously, and all identifying information was removed from step-count data before analysis, minimizing risk and protecting participant confidentiality.

Results

This section reports changes in daily step counts across three study phases for the control group (Group 1) and incentivized groups (Groups 2-4).

Table 1: Average Steps

	Phase 1	Phase 2	Phase 3
Group 1	46611.1	36964.0	32193.6
Group 2	47604.3	48720.8	40585.1
Group 3	63512.5	68874.9	59511.3
Group 4	62866.9	63907.4	47947.2

Table 1 presents average daily step counts for each group across the three phases. Baseline activity levels differed across groups, with Groups 3 and 4 exhibiting higher average step counts in Phase 1 than Groups 1 and 2. During Phase 2, all incentivized groups showed

increases in average steps, while the control group experienced a decline. By Phase 3, average step counts decreased across all groups.

Table 2: % Change in Mean Steps (vs. Phase 1 Baseline)

	Group 1	Group 2	Group 3	Group 4	All Incentives
Phase 2	-17.09%	6.35%	7.48%	3.06%	5.50%
Phase 3	-24.47%	-9.87%	-7.47%	-18.79%	-15.65%

Table 3: % Change in Median Steps (vs. Phase 1 Baseline)

	Group 1	Group 2	Group 3	Group 4	All Incentives
Phase 2	-9.53%	5.67%	16.20%	9.84%	9.84%
Phase 3	-8.75%	-8.08%	-2.95%	-4.94%	-3.13%

Tables 2 and 3 report percentage changes in mean and median steps relative to the baseline of Phase 1. In Phase 2, incentivized groups exhibited positive changes in both mean and median steps, whereas the control group showed declines. In Phase 3, step counts declined across all groups. However, the magnitude of decline was consistently smaller in the incentivized groups than in the control group, a pattern observed for both mean and median step counts.

Table 4: % Change in Mean Steps (vs. Phase 1 Baseline) Adjust Relative to Control

	Group 1	Group 2	Group 3	Group 4	All Incentives
Phase 2	0.00%	23.44%	24.57%	20.15%	22.59%
Phase 3	0.00%	14.60%	17.00%	5.68%	8.82%

Table 5: % Change in Median Steps (vs. Phase 1 Baseline) Adjust Relative to Control

	Group 1	Group 2	Group 3	Group 4	All Incentives
Phase 2	0.00%	15.20%	25.72%	19.37%	19.37%
Phase 3	0.00%	0.67%	5.80%	3.81%	5.62%

To facilitate direct comparison with the control group, percentage changes were adjusted relative to Group 1. As shown in Tables 4 and 5, incentivized groups experienced substantially less decline in Phase 3 relative to the control group. Although activity levels fell across all groups, the adjusted results indicate that incentivized groups maintained higher relative levels of physical activity, with Group 3 exhibiting the greatest attenuation of decline in both mean and median steps.

Table 6 presents statistics from the optional post-study participant survey as well as a brief summary of qualitative comments left by participants.

Table 6: Survey Results

Question 1.1: To what extent do you feel that the number of steps you recorded each week accurately reflects the number of steps you really walked? (1–5 Likert: 1 = not accurate at all, 5 = very accurate)	
Mean Rating	2.6875
Median Rating	2.5
Question 1.2: Explain your answer to the previous question.	
Several participants cited not having a phone on hand as the main reason for inaccurate step counts. Specifically, they mentioned having their phones confiscated at school or being unable to carry them while doing activities like team sports and exercise on a treadmill, noting that this led to a significant portion of steps not being counted. Some participants reported making more of an effort to carry their phones and track steps during the incentivized periods.	
Question 2.1: To what extent do you feel that the number of steps you walked each week accurately reflects your level of total physical activity? (1–5 Likert: 1 = not accurate at all, 5 = very accurate)	
Mean Rating	2.75
Median Rating	3
Question 2.2: Explain your answer to the previous question.	
Many participants reported engaging in physical activities that did not involve steps, such as cycling and weight lifting. Some participants reported prioritizing step-related activities during the incentivized periods.	
Question 3.1: If you were offered financial incentives, to what extent did they motivate you to walk more? If you were not, to what extent do you think that being offered financial incentives would have motivated you to walk more? (1–5 Likert: 1 = not at all, 5 = major motivation)	
Mean Rating	2.8125
Median Rating	3
Question 3.2: Explain your answer to the previous question.	
Responses were mixed, with some feeling highly motivated by the financial incentive, while others reported being only somewhat or not very motivated. Several participants indicated that the small size of the financial incentive was a key reason why it did not significantly increase their walking. The incentive led some participants to change their behavior specifically to ensure their steps were counted, such as carrying their phone more on the treadmill, rather than increasing their actual walking. Some participants stated that the incentive did not change their walking habits because they already had strong walking routines or because they were too busy or just forgot about the incentives.	
Question 4.1: Do you think that a much larger financial incentive—for example, \$1000—would have motivated you to take more steps than you did? (Yes/Maybe/No)	
Yes	81.25%
Maybe	18.75%
No	0%

Question 4.2: Explain your answer to the previous question.	
Most participants answered affirmatively, attributing their decision to the value of \$1000, which was described as a lot of money relative to the effort required. Some participants noted that they already walk a lot every day, so they might be able to receive the incentive without a behavioral change.	
Question 5.1: To what extent did changes in your schedule and how busy you were influence the number of steps you walked? (1–5 Likert: 1 = not at all, 5 = major influence)	
Mean Rating	3.25
Median Rating	3
Question 5.1: Explain your answer to the previous question.	
Many participants mentioned not having enough time due to being busy, having a lot of school, or busy weeks with projects, which impacted their ability to walk more. Some also noted walking more steps than usual during school vacation periods coinciding with the first phase of the study.	
Question 6. Were there any unusual circumstances (for example, an injury or illness) that might have influenced the number of steps you walked? If so, please explain	
Most participants left this answer blank or indicated that there were no unusual circumstances. Some participants mentioned events such as vacations or training for races that could have resulted in higher than average step counts for certain weeks.	

Discussion

Short-Term Effects of Financial Incentives on Physical Activity

The results suggest that financial incentives increase physical activity in the short term. As shown in Tables 2 and 3, all three incentivized groups recorded higher step counts during Phase 2 (the incentivized period) than during Phase 1 (baseline). Across the incentivized groups, this corresponded to a 5.5% increase in mean steps and a 9.84% increase in median steps during the incentive period. While these increases are smaller than those reported in some prior studies, the impact of incentives in the present study may be more dramatic than these aggregate figures suggest. Notably, the unincentivized control group (Group 1) experienced a substantial decline in activity during Phase 2, with a 17.09% decrease in mean steps and a 9.53% decrease in median steps relative to baseline. At first inspection, the magnitude of this decline in the control group may appear surprising. One possible explanation is the small sample size combined with atypical participant behavior, as several individuals reported traveling during Phase 1, which may have elevated baseline step counts. However, because such external factors may be expected to affect all groups in broadly similar ways, the relative differences between incentivized participants and the control group may be more informative than simple within-group comparisons over time. When step changes were adjusted to hold the control group constant (Tables 4 and 5), the incentivized groups demonstrated a 22.59% increase in mean steps and a 19.37% increase in median steps. This control-adjusted comparison helps isolate the effect of incentives from

broader contextual influences and suggests that incentives had a substantial short-term effect on physical activity.

Nevertheless, qualitative survey responses indicate that part of this increase may reflect changes in measurement-related behavior rather than equivalent increases in overall physical activity. Several participants reported modifying their routines to ensure steps were recorded, such as carrying their phone more consistently while using a treadmill, rather than actually increasing the amount of walking. In addition, some participants indicated prioritizing step-generating activities during the incentivized period while reducing other forms of exercise that did not contribute to step counts. Taken together, these behaviors suggest that the observed increase in steps may overstate the true impact of incentives on overall physical activity.

By contrast, other features of the study design may have worked in the opposite direction, potentially reducing the observed effect of incentives. Because participants were drawn from a convenience sample of private school students, who are likely to be more financially affluent, they may have been less responsive to small monetary incentives than lower socioeconomic-status populations. If so, the behavioral impact of similar incentives in a more economically diverse sample could be larger than what was observed here.

Long-Term (Post-Incentive) Effects of Financial Incentives on Physical Activity

The results also suggest that financial incentives may encourage physical activity in the long run, although the evidence is not as clear cut. In Phase 3, after incentives were removed, all incentivized groups recorded lower step counts than they had before incentivization, with a 15.65% decrease in mean steps and a 3.13% decrease in median steps relative to baseline of Phase 1. At first glance, the results appear consistent with the motivation crowding hypothesis, suggesting that participants who grew accustomed to financial incentives may have experienced reduced intrinsic motivation. However, as noted previously, elevated baseline activity during Phase 1 complicates this comparison. Importantly, the control group experienced an even greater decline in both mean and median steps during Phase 3 than any of the incentivized groups. When Phase 3 activity was adjusted relative to the control group (Tables 4 and 5), the incentivized groups exhibited 8.82% higher mean steps and 5.62% higher median steps than the control group. When viewed comparatively like this, the results provide tentative support for the habit formation hypothesis, suggesting that incentives may have led to physical activity that continued even beyond the incentivized period, albeit to a more limited extent.

At the same time, survey responses raise questions about whether this apparent persistence reflects durable changes in physical activity or continued sensitivity to the measurement mechanism itself. If participants learned to focus on step-tracked behaviors during the incentivized phase, some of the residual effects observed in Phase 3 may reflect changes in how activity was recorded rather than meaningful changes in physical activity. Furthermore, the study's limited time frame (two weeks per phase) may have been too short to allow them to really develop habits. If so, a longer intervention period might have allowed behavioral patterns to become more firmly established as habits, leading to a greater impact on physical activity.

Incentive Magnitude and Framing Effects

The results also shed light on how both the amount and framing of incentives may shape responsiveness. With respect to incentive magnitude, higher incentives appeared to exert a somewhat stronger influence on physical activity. This pattern is most evident when comparing Groups 2 and 3, both of which received gain-framed incentives but of different amounts (\$5 versus \$50). During the incentivized period (Phase 2), Group 3 exhibited more dramatic increases in both mean and median step counts than Group 2. Specifically, mean steps increased by 7.48% in Group 3 compared to 6.35% in Group 2, while median steps rose by 16.20% versus 5.67%, respectively. Although these differences are not extremely major, they suggest that higher incentives may cause stronger short-term behavioral responses, which is basically consistent with prior literature indicating sensitivity to incentive magnitude.

Interestingly, some differences between the two groups persisted even after incentives were removed. In Phase 3, all incentivized groups recorded lower step counts than during baseline, but the decline was less pronounced for Group 3 than for Group 2. Group 2 experienced a mean decrease of 9.87% and a median decrease of 8.08%, whereas Group 3 showed a smaller mean decrease of 7.47% and a substantially smaller median decrease of 2.95%. When changes were evaluated relative to the control group and baseline levels, Group 3 maintained larger adjusted increases than Group 2, suggesting that higher incentives may be more likely to produce residual effects after removal. From a policy perspective, however, the practical significance of these differences must be weighed against the substantially higher cost of large incentives, which may limit their feasibility despite their marginally stronger effects.

In contrast, the framing of incentives appeared to play a less decisive role. The most informative comparison in this regard is between Group 2 and Group 4, as both received \$5 incentives that were gain-framed and loss-framed, respectively. Contrary to expectations based on loss aversion theory, no consistent advantage of loss framing emerged. During the incentivized period, Group 2 demonstrated a larger increase in mean steps, while Group 4 exhibited a larger increase in median steps. A similar pattern of mixed results persisted in Phase 3, with Group 2 experiencing a smaller decline in mean steps but Group 4 showing a smaller decline in median steps. Taken together, these findings do not provide clear evidence that loss-framed incentives were more effective than gain-framed incentives in this context. One possible explanation is that participants may not have experienced the loss-framed incentive as a true loss. Although Group 4 was informed that money had been awarded and could be lost if targets were not met, participants did not physically receive the money in advance. As a result, the psychological impact of loss might not have been as strong, which in turn might have weakened the framing effect that would be expected due to loss aversion. Alternatively, it is possible that incentive framing simply plays a smaller role in fitness behavior than previous studies suggest.

Finally, qualitative survey responses point to a potential limitation of target-based incentive designs more broadly. Several participants noted that they were already exceeding the step targets prior to the intervention and therefore did not need to modify their behavior to

receive incentives. This suggests a possible ceiling effect, whereby fixed targets fail to motivate individuals who are already relatively active. In such cases, open-ended incentive structures, such as incentives proportional to total steps, might generate stronger behavioral responses, though they may also introduce greater logistical complexity and cost. While this study did not directly test such designs, these observations highlight important considerations for the design of future incentive-based fitness interventions.

Conclusion

Cumulatively, the results of this study suggest that modest financial incentives can meaningfully influence physical activity. Most significantly, and in direct response to the study's central research question, the findings indicate that some impact may persist even after incentives are removed. Although absolute step counts declined following the end of incentivization, incentivized participants maintained higher activity levels than the control group, providing tentative support for the habit formation hypothesis. While these effects are limited in magnitude, they nevertheless suggest that incentives may do more than temporarily shift behavior. From a policy perspective, this persistence is especially important. If even short-term incentives can produce residual increases in physical activity, financial incentives could represent a useful tool for public health efforts aimed at combating physical inactivity.

However, the effectiveness of such interventions depends on how incentives are designed. Comparisons across incentive conditions suggest that incentive magnitude may matter more than framing. Participants offered a higher gain-framed incentive exhibited greater short-term increases and a smaller post-incentive decline than those offered a smaller incentive, indicating the possibility of more durable effects. At the same time, these differences were relatively modest, raising questions about whether the added behavioral gains justify the substantially higher costs associated with higher incentives. Policymakers considering incentive-based interventions must therefore balance effectiveness against financial feasibility. Contrary to expectations derived from loss aversion theory, differences between gain-framed and loss-framed incentives of equal value were inconsistent and small. This suggests that framing alone may not be a decisive factor in motivating step-based activity, particularly in low-stakes or short-term contexts. In addition, survey responses indicate that target-based incentives may fail to motivate already active individuals, pointing to potential advantages of open-ended incentive designs that reward incremental effort rather than threshold achievement.

The survey data also highlight several challenges that policymakers would face in implementing incentive programs at scale. Increases in recorded steps did not always correspond to equivalent increases in actual physical activity, as participants reported carrying their phones more consistently to ensure steps were counted. Others substituted non step-based activities, such as cycling, with walking to maximize incentives. While step counts are widely used because they are easily and passively measured, these findings underscore their limitations as a comprehensive proxy for physical activity.

Given the study's small and socioeconomically disrepresentative sample, further research is needed to determine whether these findings are applicable to more diverse populations. Future studies would also benefit from longer follow-up periods, more holistic activity metrics, and direct comparisons between target-based and more open-ended incentive structures. Such work would help clarify when and under what conditions financial incentives can serve as an effective and scalable tool for promoting physical activity.

Acknowledgements

I thank the study participants for their time and diligent reporting of step-count data throughout the study. I also thank my parents for providing the financial support necessary to implement real financial incentives.

Works Cited

- Carrera, Mariana et al. "Can Financial Incentives Help People Trying to Establish New Habits? Experimental Evidence with New Gym Members." *Journal of Health Economics*, vol. 58, 2018, pp. 202–214. <https://doi.org/10.1016/j.jhealeco.2018.02.010>.
- de Buissonjé, David R. et al. "Put Your Money Where Your Feet Are: The Real-World Effects of StepBet Gamified Deposit Contracts for Physical Activity." *Internet Interventions*, vol. 31, 2023, p. 100610. <https://doi.org/10.1016/j.invent.2023.100610>.
- Elliott, Mark et al. "Physical Activity Behavior Change Driven by Engagement with an Incentivized Mobile App: Longitudinal Analysis of Daily Step Counts in Sweatcoin Users." *Journal of Medical Internet Research*, vol. 21, 2019, e12345. <https://pmc.ncbi.nlm.nih.gov/articles/PMC6643767/>.
- Jones, Damon et al. "Incentives and Habit Formation in Health Screenings: Evidence from the Illinois Workplace Wellness Study." NBER Working Paper 32745, 2024. <https://doi.org/10.3386/w32745>.
- Moller, Arlen C. et al. "Financial Motivation Undermines Potential Enjoyment in an Intensive Diet and Activity Intervention." *Journal of Behavioral Medicine*, vol. 37, 2014, pp. 819–827. <https://doi.org/10.1007/s10865-013-9542-5>.
- Patel, Mitesh S. et al. "Framing Financial Incentives to Increase Physical Activity among Overweight and Obese Adults: A Randomized, Controlled Trial." *Annals of Internal Medicine*, vol. 164, 2016, pp. 385–394. <https://doi.org/10.7326/M15-1635>.
- Salmani, Babac et al. "Financial Incentives for Physical Activity in Adults: Systematic Review and Meta-Analysis Update." *Preventive Medicine*, vol. 192, 2025, p. 108237. <https://doi.org/10.1016/j.ypmed.2025.108237>.
- Santos, Andreia Costa et al. "The Cost of Inaction on Physical Inactivity to Public Health-Care Systems: A Population-Attributable Fraction Analysis." *The Lancet Global Health*, vol. 11, 2023, pp. e32–e39. <https://pmc.ncbi.nlm.nih.gov/articles/PMC9748301/>.
- Strohacker, Kelley et al. "The Impact of Incentives on Exercise Behavior: A Systematic Review of Randomized Controlled Trials." *Annals of Behavioral Medicine*, vol. 48, 2014, pp. 92–99. <https://doi.org/10.1007/s12160-013-9577-4>.
- Strohacker, Kelley et al. "Impact of Small Monetary Incentives on Exercise in University Students." *American Journal of Health Behavior*, vol. 39, 2015, pp. 779–786. <https://doi.org/10.5993/AJHB.39.6.5>.
- StepBet. StepBet. 2026, <https://www.stepbet.com/>.
- Sweatcoin. Sweatcoin. 2026, <https://sweatco.in/>.
- World Health Organization. "Nearly 1.8 Billion Adults at Risk of Disease from Not Doing Enough Physical Activity." 26 June 2024, <https://www.who.int/news/item/26-06-2024-nearly-1.8-billion-adults-at-risk-of-disease-from-not-doing-enough-physical-activity>.
- World Health Organization. "New WHO-Led Study Says Majority of Adolescents Worldwide Are Not Sufficiently Physically Active, Putting Their Current and Future Health at Risk."

22 Nov. 2019,

<https://www.who.int/news/item/22-11-2019-new-who-led-study-says-majority-of-adolescents-worldwide-are-not-sufficiently-physically-active-putting-their-current-and-future-health-at-risk>.

Competition, Aircraft Innovation, and Airline Economics: How Manufacturer-Led Technology Change Incentivises Costs, Labour, and Performance By Zihe Wang

Abstract

This paper examines how market-driven incentives for technological innovation among commercial aircraft manufacturers are influencing airline financial performance, operational efficiency, and labour requirements over time. Drawing on industry reports, academic literature, and historical comparisons, the study iterates that intensified manufacturer rivalry, particularly between Boeing and Airbus, and more recently with the emergence of COMAC, has accelerated investment in fuel-efficient airframes, automation, and production process innovation, generating measurable downstream effects for airlines. These technological advancements are associated with reductions in unit operating costs and maintenance intensity, facilitated by fleet commonality, digital systems, and improved materials, while simultaneously reconfiguring labour demand through changes in pilot workload, training structures, and maintenance skill composition. Financial outcomes are heterogeneous across business models: low-cost carriers (LCCs) capture disproportionately larger efficiency gains from standardised fleets, whereas full-service carriers (FSCs) face more complex capital and labour trade-offs. Overall, the analysis suggests that aircraft technological innovation functions as a central economic mechanism linking manufacturer competition to airline operational restructuring and financial performance.

Keywords

Aircraft innovation; airline economics; aviation labour; fleet commonality; operational efficiency; manufacturer competition

Introduction

Commercial aircraft are among the most complex engineered products in our modern economy: combining structures, propulsion, avionics, manufacturing processes, and human workflows that together determine how air travel is priced, operated, and staffed. In recent decades, breakthroughs in composite materials, engine technology, factory automation, and digital systems have reshaped what new airframes can do: cutting fuel consumption, changing maintenance regimes, and altering the skills airlines require (International Air Transport Association [IATA], 2023; Aviation Week Network, 2022). Aircraft innovation is not only an engineering story, but an economic, organisational, and labour story whose effects touch crew rosters, maintenance shops, balance sheets, and the global flow of people and goods. These processes of innovation and decision-making ultimately cascade into the everyday operational and financial outcomes for the aviation industry.

This paper asks the following question: How do competition and market-driven incentives for technological innovation among aircraft manufacturers reshape the financial performance, operational efficiency, and labour requirements of global airlines across different

business models? Put differently, the study treats manufacturer innovation as a strategic driver of uneven operational and financial outcomes across airlines. It also acknowledges exogenous catalysts from the airlines themselves, based on the buyer's timing, fleet strategy, and business model. Existing scholarship addresses elements of this chain but rarely integrates them into a single framework.

A first strand of literature - industrial and strategic management research - explains how rivalry among manufacturers shapes product positioning and innovation incentives, including competitive-action analyses of Boeing–Airbus dynamics and strategic responses to market entry (Woo et al., 2015), as well as industry-structure treatments that situate manufacturer strategy within broader airline procurement and regulatory environments. A second strand - technical and human-factors literature - examines how automation and systems design reallocate cockpit work and introduce new supervisory and monitoring demands, including foundational models of human–automation interaction (Parasuraman et al., 2000) and empirical/operational analyses of flight-deck work organisation and performance under automation (Degani & Wiener, 1994). A third strand - airline economics and operations research- connects aircraft choice and fleet strategy to cost structures and profitability under competitive markets, including evidence on the effects of deregulation and competitive pressure on airline pricing and performance (Borenstein & Rose, 2014) and empirical findings that fleet commonality is positively associated with operating performance in low-cost carriers (Alexander, 2010). Together, these literatures establish credible mechanisms for how manufacturer incentives translate into operational labour effects and, ultimately, heterogeneous financial outcomes.

While each strand of literature offers valuable insights, they mostly operate in disciplinary silos: industrial scholars rarely quantify downstream labour effects, human-factors researchers seldom link their findings to firm-level profitability, and airline economists often treat the aircraft as an exogenous input rather than the product of strategic manufacturer behaviour.

The contribution of this paper is twofold. First, it reconceptualises the commercial aircraft as an economic intervention whose design choices shape airline operations and financial performance rather than serving merely as an engineering artefact. Second, it traces how intensified manufacturer competition and successive waves of technological innovation have restructured airline cost bases, labour deployment, and profitability over time. By linking producers' strategic innovation decisions to airlines' operational responses, the analysis clarifies how aircraft technology has historically influenced cost structures and firm-level outcomes across different phases of commercial aviation development.

The remainder of the paper proceeds as follows. Section 2 documents and analyses the incentives driving aircraft manufacturers' R&D and product strategies, with particular attention to competition, order backlogs, and process innovations. Section 3 examines the operational and labour effects of aircraft technological change, focusing on crew tasks, training dynamics, and maintenance regimes. Section 4 evaluates the financial and business-model implications for airlines, analysing how carriers capture or fail to capture the efficiency gains associated with

newer airframes. The conclusion summarises findings, discusses limitations and identification challenges, and outlines avenues for further empirical work and policy considerations.

Historical Incentives for Aircraft R&D and Technological Advancement

This section examines the economic and strategic forces that have incentivised technological research and development in commercial aircraft manufacturing over time. Rather than treating aircraft innovation as a purely technical outcome, the analysis instead embeds aircraft technological change within the competitive, demand-driven, and institutional constraints faced by manufacturers, showing that rivalry, sustained traffic growth, and persistent order backlogs have jointly shaped the scale and direction of R&D investment.

This section further contends that incentives operate not only through cost and performance considerations, but also through credibility, regulatory trust, and production feasibility, which increasingly function as economic variables in a capital-intensive and risk-sensitive industry (ICCT, 2023; KPMG, 2025). The industry's historical evolution illustrates how shifts in demand structure, regulation, and competition have shaped technological incentives over time. Early jet aircraft were developed under limited commercial demand and strong state involvement, where performance, range, and prestige dominated design priorities. Following deregulation in the 1970s and the expansion of global air travel, procurement decisions increasingly prioritised cost efficiency, scalability, and operational flexibility. Airbus's market entry intensified competitive pressure, embedding fleet commonality and pilot transition economies directly into aircraft design. In the 21st century, advances in composite materials, propulsion efficiency, and digital systems further aligned technological development with airlines' cost structures and operational constraints. The following subsections examine these mechanisms in turn, progressing from manufacturer competition to demand transformation, market signalling, and production-process innovation.

Manufacturer Competition as a Driver of Innovation

Technological advancement in commercial aircraft manufacturing has been historically and structurally driven by inter-firm competition within a highly concentrated global market. For much of the post-war period, Boeing and McDonnell Douglas dominated the commercial aviation sector, with Boeing leveraging scale, U.S. government backing, and early jet leadership to secure a commanding market position (Woo et al., 2015). However, the late 20th century introduced significant competitive disruption. Financial strain and declining commercial competitiveness at McDonnell Douglas culminated in its merger with Boeing in 1997, an event widely interpreted as both a response to and a catalyst for intensifying global competition (U.S. Department of Justice, 1997; Woo et al., 2015). Boeing's subsequent consolidation reshaped market dynamics, reducing domestic rivalry while amplifying international competition.

Airbus's emergence as a serious competitor from the 1970s onward represents one of the clearest examples of competition-induced innovation. Backed by European state investment, Airbus deliberately pursued product differentiation strategies centred on technological

integration, cockpit and fleet commonality, and aggressive pricing to overcome Boeing's incumbency advantages (Doganis, 2019). The introduction of aircraft families such as the A320 series, featuring fly-by-wire controls and common type ratings, illustrates how competitive pressure incentivised manufacturers to embed operational and labour efficiencies directly into aircraft design (Woo et al., 2015). These design choices were not technologically neutral; they were strategic responses intended to lower airlines' training and operational costs, thereby increasing Airbus's market appeal.

More recently, the entry of COMAC incentivises a newfound competition shaped by strategic state intervention and market opportunity. The development of the C919 has been framed not merely as an engineering project but as a response to perceived vulnerabilities in global supply chains and reliance on the Boeing–Airbus duopoly (U.S.-China Economic and Security Review Commission, 2020). While COMAC's current commercial impact remains limited, its entry has introduced long-term competitive pressure, particularly in the narrowbody segment, reinforcing incentives for incumbent manufacturers to accelerate R&D investment and defend market share (MIT, 2019).

Temporal Shift in Aviation Demand: From Luxury to Necessity

The nature of aviation demand has changed profoundly over time, fundamentally altering the incentives facing aircraft manufacturers. In the mid-20th century, commercial air travel was characterised by low passenger volumes, high fares, and a disproportionate share of business and government travellers (Doganis, 2019). Aircraft design priorities during this period reflected these conditions, emphasising range, speed, and prestige rather than cost efficiency or standardisation.

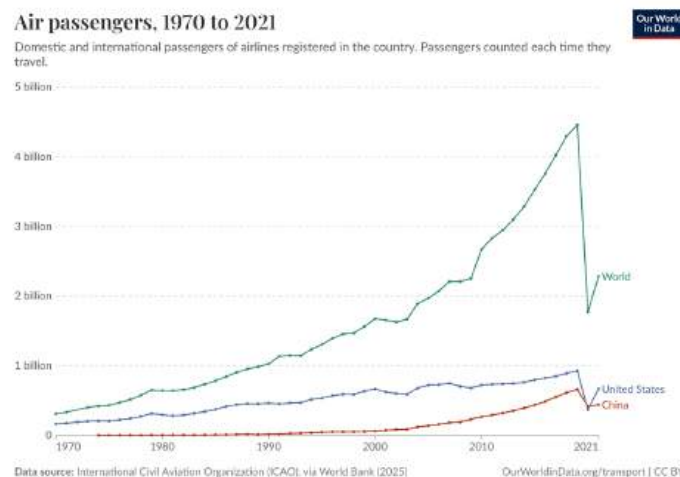


Fig 1: Air Passengers 1970-2021 (ICAO/Our World in Data) | Domestic and international passengers carried by airlines (World total vs US vs China).

From the late 1970s onward, deregulation in major markets, most notably the United States, combined with rising incomes and globalization to expand air travel to a mass consumer

base (Borenstein & Rose, 2014). Passenger numbers grew exponentially (see Figure 1): global air traffic increased from fewer than 400 million passengers annually in 1970 to over 4.5 billion prior to the COVID-19 shock (International Air Transport Association [IATA], 2019). Simultaneously, the proportion of leisure passengers increased relative to high-yield business travel, intensifying airline sensitivity to unit operating costs and fare competitiveness.

This structural demand shift altered procurement incentives. Airlines increasingly prioritised aircraft that could minimise fuel burn, maintenance downtime, and crew-related costs per seat-kilometre, rather than aircraft optimised for prestige or maximum range alone (ICCT, 2023). In response, manufacturers redirected R&D resources toward fuel-efficient engines, lightweight materials, and scalable aircraft families, aligning technological development with the economic realities of mass-market aviation.

Demand Growth for Aircraft, Order Backlogs, and Market Signaling

Empirical demand indicators further support the argument that market growth has directly incentivised technological advancement. Persistent increases in passenger demand have generated structural order backlogs, particularly in the narrowbody segment. By the late 2010s, Airbus and Boeing each reported order backlogs exceeding seven years of production at prevailing output rates (KPMG, 2025; Airbus, 2022). These backlogs (Centre for Aviation, 2026 [Figure 2]) act as powerful market signals: they provide manufacturers with revenue certainty, reduce downside risk for large capital investments, and justify long-term R&D expenditure.

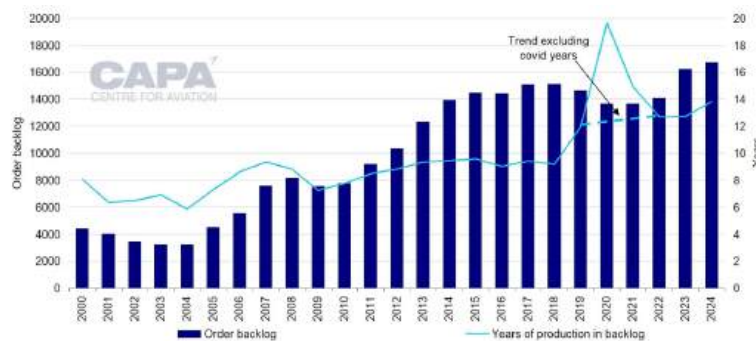


Fig 2: Bar/Line Graph showing aircraft production backlogs from 2000 to 2024 (CAPA Centre for Aviation).

The success of the A320 family relative to the Boeing 737 illustrates how demand and competition interact. Airbus's A320 series ultimately surpassed the 737 in lifetime deliveries, reflecting airlines' preference for fuel-efficient variants with high fleet commonality and lower transition costs (Reuters, 2025). Such outcomes reinforce a feedback loop in which market demand validates earlier technological choices and encourages further innovation along similar trajectories, including re-engining programs (A320neo, 737 MAX) and incremental system automation.

Beyond demand validation and cost considerations, manufacturer credibility and perceived safety performance also materially influence airline procurement decisions, reinforcing

the competitive dynamics observed between the A320 and 737 families. The grounding of the Boeing 737 MAX following two fatal accidents in 2018–2019 imposed significant reputational costs on Boeing, disrupting delivery schedules, eroding airline confidence, and altering fleet-planning timelines (Federal Aviation Administration [FAA], 2020; Joint Authorities Technical Review [JATR], 2019). Empirical reporting shows that the prolonged grounding led to order cancellations, deferred deliveries, and increased scrutiny from regulators and airline customers, weakening Boeing's competitive position in the narrowbody market during a critical period of demand growth (Reuters, 2020). Investigative findings further revealed deficiencies in certification processes, organisational safety culture, and system integration governance, amplifying concerns beyond the technical design of the aircraft itself (U.S. House Committee on Transportation and Infrastructure, 2020). In this context, airlines' preference for the A320 family cannot be attributed solely to fuel efficiency or commonality. Rather, such preferences reflect a broader assessment of manufacturer reliability, regulatory trust, and long-term operational risk, demonstrating that credibility functions as an economic variable capable of shaping demand, accelerating market share shifts, and influencing the trajectory of subsequent technological development.

Production Process Innovation and R&D Incentives

Beyond aircraft performance, competitive and demand pressures have incentivised innovation in production processes themselves. Rising order volumes and expectations of cost containment have compelled manufacturers to streamline assembly, adopt modular production architectures, and integrate digital manufacturing technologies (KPMG, 2025). The extensive use of composite materials in aircraft such as the Boeing 787 and Airbus A350 exemplifies this shift: while composites reduce weight and fuel burn, they also enable fewer parts, reduced corrosion, and longer maintenance intervals, indirectly affecting airline labour and operating costs (ICCT, 2023).

Although quantitative data on production process efficiency is less transparent than performance metrics, financial disclosures reveal sustained increases in R&D expenditure during periods of heightened competition and backlog growth (Boeing, 2021; Airbus, 2022). These investments suggest that manufacturers view process innovation as economically necessary to maintain delivery schedules, protect margins, and respond to competitive threats. Thus, it is evident that the role of market incentives are reinforced as a core driver of technological advancement.

Taken together, historical evidence indicates that technological advancement in commercial aircraft manufacturing is not an exogenous or purely engineering-driven process. Instead, it is systematically incentivised by competitive rivalry, evolving demand structures, and market signals embedded in order volumes and production constraints. These forces have shaped both the direction and timing of innovation, laying the groundwork for the operational and financial effects examined in subsequent sections.

Operational and Labour Effects of Aircraft Technological Innovation

This section examines how manufacturer-driven aircraft technological innovation reshapes airline operations by changing labour structure, skill requirements, and process efficiency. Rather than treating labour outcomes as a secondary consequence of engineering change, the analysis positions aircraft technology as an active organisational input that reallocates work between humans and systems, alters training and certification regimes, and transforms maintenance practices. By tracing developments in cockpit automation, flight-control architecture, fleet commonality, and maintenance technologies, the section demonstrates how design choices made upstream by manufacturers propagate into measurable operational and labour effects at the airline level.

Aircraft Technology and the Evolution of Flight-Crew Labour

Technological advancement in commercial aircraft has reshaped the structure, scale, and skill composition of flight-crew labour, operating through both technical and economic channels. In the early jet age, cockpit labour requirements were high due to limited automation and system integration. Aircraft such as the Boeing 707 and early wide-body jets required a three-person cockpit, including a flight engineer tasked with continuous monitoring and manual management of propulsion, electrical, and environmental systems (Degani & Wiener, 1994). These labour requirements reflected technological constraints rather than operational necessity.

Progressive advances in avionics, full authority digital engine control units (FADEC), and integrated aircraft systems enabled the automation of tasks previously performed manually by flight engineers. By the late twentieth century, these developments led to regulatory approval of two-pilot cockpits across virtually all commercial aircraft types (Amalberti, 2017). This transition represents a major labour-saving effect associated with manufacturer innovation, reducing crew staffing requirements and lowering crew costs per flight hour.

However, research in aviation human factors consistently emphasises that automation does not eliminate labour, but reconfigures it. Parasuraman et al. (2000) demonstrate that increasing levels of automation shift pilots' roles from active controllers to system supervisors, altering workload distribution rather than reducing total cognitive demand. Empirical simulator and line-operation studies show that while routine phases of flight require less manual input, abnormal situations demand rapid re-engagement and high-level decision-making, increasing the importance of training, procedural discipline, and experience (Wickens et al., 2015). As a result, technological advancement reduces the quantity of cockpit labour while increasing its skill intensity, reinforcing airlines' reliance on highly trained pilots rather than permitting wholesale labour substitution.

Fly-by-Wire, Systems Integration, and Operational Efficiency

Fly-by-wire (FBW) technology exemplifies how manufacturer design choices translate into airline operational efficiency. FBW systems replace mechanical linkages with digital control laws that mediate pilot inputs, enforcing flight-envelope protections and optimising control

responses. First deployed commercially on the Airbus A320, FBW has since become standard across most new-generation aircraft (Airbus, 2007).

From an operational standpoint, FBW contributes to efficiency in several tangible ways. First, it standardises aircraft handling characteristics across variants, reducing performance variability between aircraft types and simplifying crew deployment (Amalberti, 2017). Second, envelope protection reduces the likelihood of exceedances that lead to maintenance inspections, aircraft-on-ground events, or insurance claims, indirectly lowering operational disruption costs (Degani, 2004). Third, FBW facilitates system redundancy and fault isolation, improving dispatch reliability, an operational metric closely linked to airline revenue protection and schedule integrity (IATA, 2023).

Empirical safety and performance studies show that FBW-equipped aircraft can reduce loss-of-control risk in normal operations, although the technology also introduces risks linked to mode awareness and automation dependency. In this sense, a balanced interpretation is necessary, as the same digital mediation that enhances safety and operational consistency can also introduce new forms of vulnerability (Wickens et al., 2015). Research in human factors documents risks associated with automation dependency, including skill degradation in manual flight operations, reduced situation awareness, and mode confusion under abnormal conditions. In highly automated cockpits, pilots may engage less frequently in direct control inputs, potentially diminishing hands-on proficiency over time. Moreover, overreliance on automated systems can delay intervention when system behaviour deviates from expectations, thereby transferring operational risk rather than eliminating it.

Accordingly, the efficiency gains derived from fly-by-wire and advanced automation are not autonomous technological outcomes; but instead are contingent upon effective training regimes, procedural discipline, and organisational oversight. Automation enhances performance when embedded within a human-centred system design that preserves pilot authority and cognitive engagement. This duality reinforces the broader argument that technological innovation in aviation reconfigures labour rather than displacing it, embedding new supervisory demands within the operational structure while retaining efficiency. Thus, these findings underscore that FBW enhances efficiency by embedding operational discipline into aircraft design, rather than by removing pilots from the loop. In doing so, it helps reduce costly deviations and operational inefficiencies.

Training Accessibility, Type Rating, and Fleet Commonality

Training constitutes a major fixed labour cost for airlines, making training accessibility a central economic consideration in aircraft procurement decisions. Manufacturers increasingly design aircraft families with high levels of cockpit and systems commonality, enabling pilots to transition between variants with minimal additional training. Airbus's cross-crew qualification (CCQ) framework allows pilots to operate multiple aircraft types within a family after limited differences training, significantly reducing simulator hours and instructor costs (Airbus, 2022).

Regulatory data support these efficiency gains. Certification standards published by

EASA and the FAA specify reduced transition training requirements for aircraft sharing common type ratings, compared to mixed-manufacturer or legacy fleets (EASA, 2021). Academic research confirms that commonality lowers both initial and recurrent training costs, improves crew scheduling flexibility, and reduces the administrative burden associated with maintaining multiple pilot pools (Doganis, 2019).

These effects are particularly pronounced for low-cost carriers, whose business models rely on high aircraft utilisation and tight cost control. Empirical studies of airline operations show that training cost savings scale with fleet size and homogeneity, creating strong incentives for airlines to concentrate orders within a single aircraft family (Graham, 2013). Thus, training accessibility functions as a labour-efficiency mechanism increasingly embedded by manufacturers to increase the attractiveness of their products.

Maintenance, Repair, and Overhaul (MRO): Labour and Process Effects

Technological innovation has also reshaped airline labour through its impact on maintenance, repair, and overhaul (MRO) activities. Advances in materials science, component modularity, and onboard diagnostics have altered both the frequency and nature of maintenance work. Composite airframes exhibit greater resistance to corrosion and fatigue cracking than aluminium structures, extending inspection intervals and reducing labour-intensive heavy checks (ICCT, 2023). While composite repair requires specialised skills, studies show that overall labour hours per flight cycle decline as fleets mature (IATA, 2023).

Digital health-monitoring systems represent another major shift. Modern aircraft continuously collect performance data, enabling predictive maintenance regimes that identify component degradation before failure occurs. This reduces unscheduled maintenance events, which are labour-intensive and operationally disruptive, and reallocates labour toward planned, data-driven interventions (IATA, 2023). From a labour economics perspective, this transition raises productivity by smoothing labour demand and reducing costly overtime and AOG-related inefficiencies.

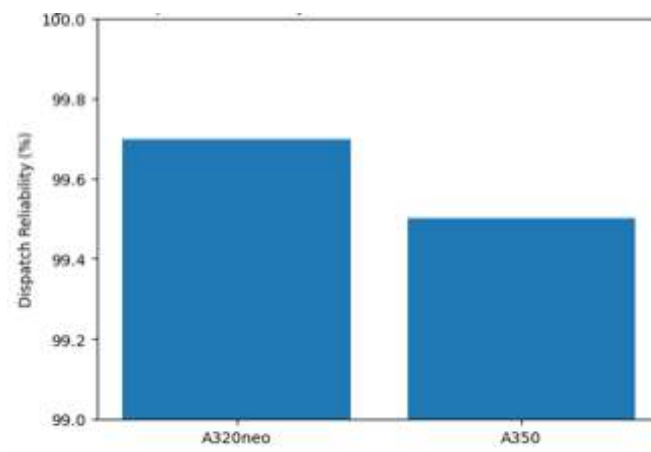


Fig 3: Dispatch reliability of Selected New Generation Airbus (Original figure, data derived from Airbus, 2023/24).

As shown in Figure 3, manufacturer-reported dispatch reliability rates for new-generation aircraft are approximately 99.5–99.7%. Because dispatch reliability reflects departures free from technical delay, it serves as a proxy for unscheduled maintenance frequency. Higher reliability reduces AOG events and reactive maintenance, stabilising technician workload and improving labour productivity (Airbus, 2023; 2024).

Fleet commonality further amplifies these gains. Airlines operating single-type or single-manufacturer fleets benefit from standardised tooling, simplified technician training, and reduced spare-parts inventories (Doganis, 2019). Southwest Airlines' exclusive operation of Boeing 737 aircraft has been extensively documented as a source of maintenance efficiency, with lower maintenance cost per available seat kilometre relative to mixed-fleet carriers (Gittell, 2005). In contrast, legacy airlines operating complex fleets incur higher fixed labour and inventory costs, illustrating how aircraft technology and procurement strategy jointly shape MRO labour requirements.

Taken together, the evidence demonstrates that aircraft technological innovation reshapes airline labour and operations through a set of interrelated mechanisms. Automation and systems integration reduce routine manual workload while increasing the skill intensity of cockpit labour. Fly-by-wire and digital systems enhance operational efficiency by stabilising performance and reducing costly disruptions. Training commonality lowers labour fixed costs and increases scheduling flexibility, while advances in materials and diagnostics transform maintenance labour from reactive to predictive. These outcomes are neither accidental nor purely technological; they reflect manufacturer strategies that embed labour and operational efficiencies into aircraft design, reinforcing the competitive link between innovation, airline operations, and economic performance.

Financial and Market Effects on Airline Firms

This section examines how aircraft technological innovation, driven by manufacturer competition and market incentives, propagates beyond operations and labour to shape airline-level financial performance and market outcomes. Treating aircraft technology as a strategic asset, the section shows how efficiency gains affect cost structures, asset productivity, revenue resilience, and competitive positioning. While newer aircraft systematically reduce unit operating costs, their financial benefits are unevenly distributed across airlines, reflecting differences in demand exposure, capital structure, and corporate strategy. By situating technological adoption within demand cycles, macroeconomic shocks, and heterogeneous business models, the section analyses competition and demand effects on airline revenues and costs, the role of exogenous disruptions in financial resilience, and how low-cost and full-service carriers leverage aircraft technology in distinct ways to generate divergent financial outcomes.

Competition, Demand Growth, and Unit Cost Economics

Airline financial performance is shaped by the interaction between passenger demand growth and cost competition, both of which amplify the economic value of technological

innovation in aircraft. As global passenger demand expanded rapidly in the decades preceding the COVID-19 pandemic, airlines faced intensifying price competition, particularly in liberalised markets (Doganis, 2019). Economic theory and empirical evidence indicate that in such environments, cost reductions achieved through improved fuel efficiency, higher aircraft utilisation, and lower maintenance intensity translate into competitive advantage (Borenstein & Rose, 2014).

Aircraft technology plays a central role in this mechanism. Newer airframes offer lower fuel burn per seat-kilometre and improved reliability, reducing variable costs that account for a substantial share of airline operating expenses (ICCT, 2023). Next-generation narrowbody aircraft such as the Airbus A320neo and Boeing 737 MAX deliver approximately 15–20% lower fuel burn per seat relative to prior-generation variants (ICCT, 2023), while wide-body aircraft such as the Boeing 787 and Airbus A350 achieve fuel efficiency improvements of roughly 20–25% compared to legacy counterparts. Given that fuel accounts for 20–30% of airline operating costs, such efficiency gains translate into structural reductions in cost per available seat kilometre (CASK), particularly under high-utilisation conditions.

In addition, IATA data show that fuel costs alone have historically represented 20–30% of airline operating costs, making even marginal efficiency gains financially significant (IATA, 2019). As a result, airlines operating more technologically advanced fleets are better positioned to sustain profitability under fare pressure, whereas carriers with older fleets face cost disadvantages that erode margins during competitive cycles.

The financial implications of this dynamic are visible in fleet renewal patterns. Airlines consistently allocate significant capital expenditure toward next-generation aircraft during periods of sustained demand growth, indicating that aircraft technology is treated as a strategic financial asset rather than a passive input (Airbus, 2022; Boeing, 2021). This reinforces the broader thesis that manufacturer-driven technological innovation becomes financially consequential only insofar as it interacts with airline demand conditions and competitive intensity.

Market Shocks, Exogenous Disruptions, and Financial Resilience

While technological efficiency improves baseline financial performance, airline profitability remains highly sensitive to exogenous market shocks, including pandemics, geopolitical conflict, and regulatory disruptions. The COVID-19 pandemic provides a particularly stark illustration: global passenger traffic declined by over 60% in 2020, producing unprecedented revenue losses across the industry.

Specifically, the global airline industry recorded net losses exceeding USD 120 billion (IATA, 2021), with industry debt levels increasing by more than USD 200 billion as carriers sought liquidity. In such an environment, marginal differences in fuel efficiency and maintenance intensity may materially influence cash burn rates and survival probabilities. In this context, aircraft technology did not prevent financial distress, but it materially affected airlines' capacity to absorb and recover from the shock.

Industry guidance issued during the pandemic indicates that airlines with newer, more fuel-efficient fleets were able to ground aircraft more flexibly (due to lower storage and technical reactivation requirements allowing for lower grounded fixed costs), reduce operating cash burn, and resume services more rapidly as demand returned (ICAO, 2022). Similarly, aircraft with longer range and lower trip costs enabled airlines to reconfigure networks in response to border closures and shifting demand patterns. These outcomes suggest that technological innovation enhances financial resilience, even when short-term demand collapses overwhelm normal cost advantages.

Geopolitical disruptions, such as fuel price volatility linked to conflict, further underscore this relationship. Economic studies show that airlines with lower fuel intensity experience less margin compression during fuel price shocks, stabilising cash flows relative to less efficient competitors (Borenstein, 2011). Thus, while aircraft technology cannot insulate airlines from macroeconomic shocks, it meaningfully conditions their financial exposure and recovery trajectories.

Business Model Heterogeneity and Strategic Use of Technology

The financial effects of aircraft innovation vary systematically across airline business models. Low-cost carriers (LCCs) and full-service network carriers (FSCs) pursue different corporate strategies, shaping how they extract value from technological advancement. LCCs prioritise cost minimisation and asset utilisation, making fuel efficiency, fleet commonality, and simplified maintenance central to their profitability (Gittell, 2005; Doganis, 2019). Empirical evidence shows that LCCs operating homogeneous fleets achieve lower unit costs and higher margins, partly due to reduced training, maintenance, and scheduling complexity (Graham, 2013).

In contrast, FSCs emphasise network connectivity, service differentiation, and premium cabin revenue. For these carriers, technological innovation supports financial performance through enhanced range, payload flexibility, and passenger experience rather than pure cost reduction (Doganis, 2019). Wide-body aircraft innovations, such as composite structures and advanced engines, enable long-haul route economics that would be infeasible with older technology, supporting revenue diversification even at higher capital cost.

This divergence illustrates that aircraft technology does not produce uniform financial outcomes. Instead, its value is mediated by corporate strategy and market positioning, reinforcing the paper's argument that manufacturer innovation interacts with airline-level decision-making rather than determining outcomes unilaterally.

Counterargument: Capital Intensity and Financial Risk

A potential counterargument is that technological innovation increases financial risk by raising capital intensity and debt exposure. New aircraft involve substantial upfront investment, and empirical studies note that aggressive fleet renewal can weaken balance sheets if demand forecasts prove overly optimistic (Gillen & Lall, 2004). The financial difficulties faced by some

airlines following rapid fleet expansion prior to economic downturns support this concern.

However, this argument overlooks the dynamic nature of aircraft finance and residual value. Leasing markets can reduce upfront risk while sometimes increasing ongoing cost and reducing flexibility. Likewise, manufacturer financing and secondary aircraft markets mitigate capital risk by spreading exposure over time (Doganis, 2019). Moreover, studies indicate that technologically advanced aircraft retain higher residual values, reducing long-term financial risk relative to older models (ICAO, 2022). Consequently, while technological adoption entails financial risk, the literature suggests that these risks are conditional on managerial strategy rather than inherent to innovation itself.

This section demonstrates that aircraft technological innovation reshapes airline financial performance through multiple channels: reducing unit operating costs, enhancing resilience to market shocks, and enabling differentiated corporate strategies. Demand growth and competition amplify the value of efficiency gains, while exogenous disruptions reveal the importance of technological flexibility in financial recovery. Crucially, the financial effects of innovation are heterogeneous, reflecting differences in business models, capital structures, and strategic priorities. These findings reinforce the paper's central claim that aircraft technology functions as a critical economic lever linking manufacturer competition to airline-level financial outcomes.

Conclusion

This study argues that advances in commercial aircraft technology emerge from economic and strategic pressures rather than from engineering trajectories alone. Across successive competitive regimes, spanning the Boeing–McDonnell Douglas era, Airbus's rise, and the entry of COMAC, manufacturers have responded to rivalry and market conditions by directing innovation toward materials, automation, systems integration, and production efficiency. These manufacturer-level decisions have propagated through airline operations, reshaping labour organisation and cost structures and producing financially differentiated outcomes across business models and time periods.

Sections 2–4 collectively show that technological change in commercial aviation is not exogenous, but endogenously shaped by competitive pressures, sustained passenger growth, and institutional constraints, with effects that propagate through operations into financial performance. Persistent demand expansion and order backlogs incentivise manufacturers to prioritise fuel efficiency, reliability, fleet commonality, and scalable design, while competition shapes both the pace and direction of innovation. Episodes such as the 737 MAX crisis further demonstrate that credibility and regulatory trust function as binding economic constraints on technological development. These design choices restructure airline production functions by embedding labour efficiency into aircraft architecture. Through automation, training commonality, and predictive maintenance, routine workload is reduced while skill intensity is increased. The resulting cost and productivity gains are real but unevenly captured, depending on demand exposure, capital structure, and corporate strategy: low-cost carriers systematically monetise standardisation, whereas full-service carriers deploy technology to sustain network

complexity at higher capital intensity. Ultimately, aircraft innovation enhances resilience, improving airlines' capacity to absorb shocks such as pandemics and fuel price volatility without eliminating exposure to cyclical risk.

A key limitation of this study is its primarily synthetic and interpretive design. The analysis integrates historical evidence, industry reports, and secondary academic literature to construct a coherent explanatory framework; however, its causal claims are inferred rather than formally identified through primary data collection or a dedicated econometric research design. While triangulation across mixed sources strengthens conceptual coherence and internal consistency, the absence of direct quantitative identification limits the ability to isolate the independent effect of aircraft technology on labour performance and corresponding financial outcomes.

Several avenues for further research follow. Future work could examine how tightening environmental regulation and alternative propulsion technologies reshape innovation incentives and airline economics, particularly as sustainable aviation fuels and non-conventional aircraft mature. Methodologically, panel studies exploiting staggered fleet introductions or regulatory shocks would strengthen causal identification of technology's labour and financial effects.

Overall, the paper reframes the commercial aircraft as an economic intervention rather than a neutral engineering artefact, showing that design choices made upstream by manufacturers have lasting consequences for airline cost, labour organisation, and competitive performance. By linking manufacturer competition to airline labour and financial outcomes through technological design choices, it clarifies how innovation reshapes the structure, performance, and resilience of the global aviation industry over time.

Acknowledgments

I would like to thank Dr Georgios Aivaliotis, who was my initial mentor in developing this paper. I would also like to thank Dr George Zifkos, for his editorial guidance and assistance with final revisions.

Works Cited

- Alexander, B. (2010). How fleet commonality influences low-cost airline operating performance: Empirical evidence. *Journal of Air Transport Management*, 16(6), 299–303. <https://www.sciencedirect.com/science/article/abs/pii/S0969699710000293>
- Airbus. (2007). Fly-by-wire: A320 family technical background. Airbus S.A.S.
- Airbus. (2022). Annual report 2022. Airbus SE.
- Airbus. (2023). A320neo family: Facts and figures. Airbus S.A.S.
- Airbus. (2024). A350 XWB: Facts and figures. Airbus S.A.S.
- Aviation Week Network. (2022). How next-generation aircraft are reshaping airline maintenance and training. *Aviation Week & Space Technology*. <https://aviationweek.com>
- Boeing. (2021). Annual report 2021. The Boeing Company.
- Borenstein, S. (2011). On the implications of oil price shocks for airline profitability. *Energy Journal*, 32(1), 1–26.
- Borenstein, S., & Rose, N. L. (2014). How airline markets work... or do they? Regulatory reform in the airline industry. *Journal of Economic Perspectives*, 28(2), 3–24. <https://doi.org/10.1257/jep.28.2.3>
- Centre for Aviation. (2026). Aircraft order backlog eases to 12 years as Airbus outsells Boeing again in 2025 <https://centreforaviation.com/analysis/reports/global-aircraft-order-backlog-another-new-record-at-14-years-airbus-outsells-boeing-again-707386>
- Centre for Aviation. (2025). Global aircraft order backlog: Another new record at 14 years as Airbus outsells Boeing again. <https://centreforaviation.com/analysis/reports/global-aircraft-order-backlog-another-new-record-at-14-years-airbus-outsells-boeing-again-707386>
- Degani, A. (2004). Modes in human-machine systems: Review, classification, and application. NASA Ames Research Center.
- Degani, A., & Wiener, E. L. (1994). On the design of flight-deck procedures. NASA Contractor Report.
- Doganis, R. (2019). *Flying off course: Airline economics and marketing* (5th ed.). Routledge.
- European Union Aviation Safety Agency. (2021). Type rating and transition training requirements. EASA.
- Federal Aviation Administration. (2020). Boeing 737 MAX return to service. U.S. Department of Transportation.
- Gillen, D., & Lall, A. (2004). Competitive advantage of low-cost carriers: Some implications for airports. *Journal of Air Transport Management*, 10(1), 41–50.
- Gittell, J. H. (2005). *The Southwest Airlines way*. McGraw-Hill.
- Graham, A. (2013). *Managing airports: An international perspective* (4th ed.). Routledge.
- International Air Transport Association. (2019). Air passenger market analysis. IATA.
- International Air Transport Association. (2021). World air transport statistics. IATA.

- International Air Transport Association. (2023). Economic performance of the airline industry. IATA.
- International Civil Aviation Organization. (2022). Effects of COVID-19 on civil aviation: Economic impact analysis. ICAO.
- International Council on Clean Transportation. (2023). Cost assessment of near- and mid-term technologies to improve new aircraft fuel efficiency. ICCT.
- Joint Authorities Technical Review. (2019). Boeing 737 MAX flight control system: Observations, findings, and recommendations.
- KPMG. (2025). Emerging trends in aerospace and defense. KPMG International.
- MIT. (2019). China's COMAC and the challenge to the Airbus–Boeing duopoly (Master's thesis). Massachusetts Institute of Technology.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 30(3), 286–297.
<https://doi.org/10.1109/3468.844354>
- Reuters. (2025). Airbus A320 overtakes Boeing 737 in lifetime deliveries. Reuters.
<https://www.reuters.com/business/aerospace-defense/airbus-a320-flies-past-boeing-737-most-delivered-jet-history-2025-10-07/#:~:text=Boeing's%20decades%2Dold%20record%20fell,based%20aviation%20analytics%20firm%20Cirium.>
- Reuters. (2020). Boeing faces order cancellations and delivery delays amid the 737 MAX crisis. Reuters.
<https://www.reuters.com/article/world/boeing-737-max-cancellations-top-350-planes-in-first-half-of-2020-idUSKCN24F2BH/>
- Reuters. (2022). China certifies C919 seeks to compete with Boeing and Airbus globally. Reuters.
<https://www.reuters.com/business/aerospace-defense/china-certifies-c919-jet-compete-with-airbus-boeing-photos-2022-09-29/>
- Reuters. (2023). Airbus and Boeing face record backlogs as airlines race to renew fleets. Reuters.
<https://www.reuters.com/business/aerospace-defense/airbus-posts-record-orders-keeps-delivery-crown-2023-2024-01-11/>
- U.S. Department of Justice. (1997). *United States v. Boeing Co. and McDonnell Douglas Corp.: Competitive impact statement*.
- U.S. House Committee on Transportation and Infrastructure. (2020). *The design, development & certification of the Boeing 737 MAX*. U.S. House of Representatives.
- U.S.-China Economic and Security Review Commission. (2020). *Commercial aircraft development in China*. U.S. Government Publishing Office.
- Woo, J., Park, Y., & Sung, T. (2015). An analysis of the competitive actions of Boeing and Airbus in the aerospace industry. *Research Policy*, 44(2), 437–451.
- Wickens, C. D., Lee, J., Liu, Y., & Gordon-Becker, S. (2015). *An introduction to human factors engineering* (2nd ed.). Pearson.

World Bank. (2025). Air transport, passengers carried (ICAO data) [Dataset]. World Development Indicators. <https://data.worldbank.org/indicator/IS.AIR.PSGR>
Author

Zihe Wang is a Year 12 student in Melbourne and a licensed pilot whose academic interests centre on aviation economics and operational structure. He intends to study finance and pursue aviation management post-studies. In this paper, he examines how technological innovation interacts with labour markets and airline financial performance.

Greek Literary Education and Rhetorical Training in the Poetic Inscription of Q. Sulpicius Maximus By Meilin Sha

Abstract

This paper examines the work of the boy poet Q. Sulpicius Maximus as a product of the imperial Roman education system. Sulpicius' poem was written in Greek from the perspective of Zeus scolding Helios for the misconduct of Phaethon, and he delivered it extemporaneously in a public competition. Sulpicius' education reflects the Homeric tradition and the Greek model that Roman rhetorical training was based upon. This paper illustrates different adaptations of Greek material by analyzing the Sulpicius' poem in comparison to the works of Ovid and Lucian, who also wrote about Phaethon. Surviving student compositions on papyri from Greco-Roman Egypt also reinforce the rarity of Sulpicius' narrative, as his version exemplifies extensive elite Roman training from within the city of Rome itself. Through close examination of Greek and Roman literary models, this study argues that Sulpicius' poem provides a broader lens into the extent of which Greek culture influenced Roman schooling, which allowed students to produce impressive works of literature and rhetoric.

Introduction

In 1871, excavations on the Via Salaria led to the discovery of the tombstone of Quintus Sulpicius Maximus. The marble slab is covered in inscriptions, including an epitaph in Latin and a poem in Greek. A closer look at the Latin inscription reveals that Sulpicius was only a boy of 11 years, 5 months, and 12 days at the time of his death in 94 CE. His parents, Quintus Sulpicius Eugramus and Licinia Januaria, state that they erected the tombstone to both mourn their son and to preserve his poetic talent for posterity: Consisting of 43 lines in Greek, the youth's poem is written in the perspective of Zeus scolding Helios for lending his chariot to Phaethon. Both the style and the subject matter of Sulpicius' poem display evident signs of Greek cultural influence.

Sulpicius lived during the early empire under the rule of Domitian, who instituted Greek-style poetry contests dedicated to Jupiter Capitolinus. Although the boy possessed a Roman cognomen and belonged to the Claudian gens, signalling his status as a citizen, his father was most likely a freedman, since he had the Greek cognomen Eugramus. As was the tradition, slaves would often take the praenomen and nomen of their former masters after being freed. Prior to his death, Sulpicius participated in an extemporaneous contest between 52 Greek poets as part of the third series of the Capitoline games. His parents proudly state that he gained distinction and public favor because of his talent at such a young age, prompting them to inscribe his poem on the tombstone.¹ I will analyze his poem carefully in the context of Roman literary and rhetorical education, demonstrating the foundational role of Greek models in shaping Sulpicius' education. I will then examine Sulpicius' possible sources of inspiration for the Phaethon story itself. The paper ends with a section that compares Sulpicius' poem to the poetic compositions of students discovered on papyri in Egypt.

Introduction to Roman Rhetorical Education²

The earliest stages of Roman education began not with Latin, but Greek. Quintilian tells us in his *Institutio Oratoria* “that Latin learning is derived from Greek is a further reason for [a student’s] being first instructed in the latter.”³ This model of Greek education, known as the *enkyklios paideia*, referenced the complete fundamentals of education. The first stage consisted of basic proficiency in reading, writing, and arithmetic. After mastering those primary skills, students studied under a grammarian, where they were introduced to literary texts, starting with the poetry of Homer. Only then did some elite young men proceed to schools of rhetoric where they read works in prose and were trained in oral expression.

This Greek-centered style of education started to gain prominence during the 1st Century CE, when the scholars of the Second Sophistic intellectual movement revived the philosophy and learning of ancient Greece. Sophisticated literacy in Greek had previously been confined only to a small group of aristocrats, but now it grew to encompass a broader spectrum of inhabitants of the Roman empire, providing opportunities for social advancement to those from undistinguished backgrounds. New styles of Greek writing flourished: the rhetorical style of Atticism, for example, emerged because of Roman fascination with the Athenian past, as scholars aimed to purify literature by returning to standards of the Greek classical period.⁴ This fascination with Greek learning transferred to the education of young Roman pupils, as they were granted the advantage of bilingualism by being introduced to literature with the works of Homer. Many fundamental exercises involved copying passages from specific excerpts, and students composed *scholia minora*, which were commentaries that included words directly from the Homeric texts.⁵

Once children mastered basic literacy in both Greek and Latin, they would then choose to become orators or philosophers. Greek rhetorical training emphasized public delivery, progressing from preliminary exercises, *progymnasmata*, to whole declamations, *meletai*. Rhetorical training rested on the framework of poetry, as *progymnasmata* were often based on Homeric lines. Roman education adapted these exercises through the practice of declamation, which included the forms of *controversiae* and *suasoriae*. In *controversia*, students would argue in a mock trial about legal questions in Greek and Roman law, while *suasoria* required an orator to attempt to guide the action of a famous historical or mythical person. Rhetorical training also involved exercises in praise (*encomium*) and impersonation (*êthopoia*), the latter of which is exemplified in the poem of Q. Sulpicius Maximus, who assumes the persona of Zeus, berating Helius.⁶ Indeed, papyri found in Egypt preserve examples of these two specific exercises, attesting to their prevalence in the sphere of Roman education across the empire. Raffaella Criboire notes that most papyri contained exercises written in poetry rather than prose, as “rhetorical training not only was based upon previous knowledge of the poets but also reinforced and expanded students’ poetic skills.”⁷

Rhetorical development was closely tied to competition and public performance. The Capitoline Games that Q. Sulpicius Maximus participated in were not the only ones of their kind, as evidence from Egypt demonstrates that similar games in the cities of Oxyrhynchos and Anchinoë, on the far-flung borders of the empire, also included contests in poetry.⁸ Competition

remained critical throughout a Roman boy's education and professional development, as professional rhetors displayed their skill by arguing in public orations. Students would often attend these display speeches (*epideixis*) to audit the skill of popular sophists and particular teachers. Quintilian accentuates the role of competition in molding young students: "Sometimes, if he is recalcitrant, let another receive his lesson so that he may envy him; and let him occasionally be in competition, and feel more often than not that he is the winner. He should even be enticed by the prizes that boys of that age strive for."⁹ Taken together, these educational practices in the Roman empire explain how an eleven-year-old boy could compose a sophisticated poem in Greek that drew confidently on Homeric language and mythological content.

Q. Sulpicius Maximus's Poem as a Work of Imperial Roman Education

While Sulpicius' poem contains language that mirrors that of Homer, the story of Phaethon does not appear in the *Iliad* or *Odyssey*, but is attested in a variety of other Greek authors like Euripides and Apollonius of Rhodes, with Hyginus dating the story back to Hesiod.¹⁰ While Euripides wrote a tragedy centering on Phaethon that is now lost, he alludes to the myth in his play *Hippolytus*, mentioning the Eridanus river as the site where Phaethon crashed.¹¹ Apollonius expands on the Eridanus, noting that at the time of the Argonauts, it was still smoldering from Phaethon's ashes after he was smitten by a thunderbolt. He also mentions the transformation of Phaethon's sisters into poplar trees that shed tears of amber along the riverbank.¹²

Perhaps the most detailed account of Phaethon's story belongs to Books I and II of Ovid's *Metamorphoses*, which are a likely source for Sulpicius, as Ovid slightly preceded Sulpicius and was widely known in antiquity—at the very least, Ovid's version can be taken to represent a version of the myth in current circulation in the early empire. (Table 1 presents a comparison between Sulpicius' Phaethon story, and the versions of Ovid and Lucian.) Ovid describes Phaethon's struggle with his parentage, as he is mocked by Epaphus after claiming to be the son of Phoebus Apollo. In an effort to prove his divine lineage and his contention that Apollo will grant any of his wishes, Phaethon asks to drive Apollo's chariot. While Ovid is the first to place emphasis on Phaethon's parentage, he also draws inspiration from previous Greek authors, sharing Euripides's portrayal of Clymene, Phaethon's mother, whose warnings are in vain.¹³ Notably, Ovid features Tellus (earth) in his retelling of the Phaethon story. Sulpicius' poem also describes how the earth was harmed by the scorching flames, but—unlike Ovid—this features in Zeus' direct speech, rather than the narrator's description of Phaethon's disastrous ride: it is tempting to infer that Sulpicius is here reworking Ovid's story to align more closely with Greek models that might have featured extended direct dialogue between Zeus and Helios. (There is no exchange between Juppiter and Sol/Helios in the *Metamorphoses* version.) Although the works of both Sulpicius and Ovid are rooted in traditions of the Phaethon myth, Sulpicius' choice to compose his poem in Greek aligns his work more closely with Greek epic conventions, revealing

a stronger connection to the style of his early education and the strong influence of elite Greek culture at Rome

Another poet who depicted the story of Phaethon was Lucian of Samasota, who lived in the 2nd century CE. Although he was Syrian by birth, Lucian composed satire in Attic Greek, deliberately exaggerating Homeric conceptions of the gods. His retelling of the Phaethon myth is strikingly similar to that of Sulpicius', as Lucian also writes from the perspective of Zeus, reprimanding Helios in direct speech.¹⁴ In both Lucian and Sulpicius' versions, Zeus threatens to use his thunderbolts to punish Helios. This shared narrative of Homeric Zeus indicates a broader return to previous Greek models, distinguishing their works from Ovid's version. Lucian's ability to purposefully manipulate Homeric convention and language demonstrates a strong foundation in Homeric style. Because Sulpicius' *ex tempore* composition also follows a similar model, we can assume that both authors relied on audiences familiar with Greek models and language. The prominence of bilingualism and the major role Greek poetry played in early education also demonstrate the influence of the Greek model of study across the vast expanse of the empire. By contrast, although Ovid engages with topics of Greek mythology, the earlier period in which he lived was before the Second Sophistic left a lasting legacy of Hellenization. Only later in the imperial period did Greek become a hallmark for rhetorical study, as reflected in the works of Lucian and Sulpicius. This later boom in widespread Greek bilingualism functioned as a cultural tie, bringing audiences together with a unified literary tradition.

Sulpicius' composition shows influence of the rhetorical exercise of *êthopoia*, since he impersonates Homeric Zeus and Helios in his dialogue. By using Homeric stylistic choices like epithets, Sulpicius reflects his early training and internalization of the poetry of Homer, since Roman rhetorical training encouraged such references to the Greek world. Not only was poetry used as a tool to instruct students in exercises of grammar, but it gave students like Sulpicius a repertoire of terms and a toolkit of rhetorical skills. The public and competitive aspect of delivering his poem to an audience likely motivated Sulpicius to both display creativity but also prove that he was well-trained in the conventional education model. This paved the foundation for rhetorical standards in the empire, since Greek literary traditions were popular among later authors. Throughout Greco-Roman history, each author's reinvention of the same topic of Phaethon demonstrates that Roman education encouraged extemporaneous creativity and different interpretations. The foundation of Greek epic models, reveals the enduring authority of Homeric tradition within imperial education. Sulpicius' poem thus stands not merely as a literary work, but as evidence of how Greek culture continued to shape Roman intellectual identity throughout the course of the empire.

Table 1: Comparison of Versions of the Phaethon Story—Sulpicius, Ovid, Lucian

Sulpicius	Ovid	Lucian
"All the earth groaned on account of the malignant boy; and at that time I extinguished the	Earth spoke. Then, unable any longer to endure the heat or keep on talking, she pulled her face back	—

impressive fire.” ¹⁵	deep within herself, into caves closer to the dead below.” ¹⁶	
“But if he has a different thought of quitting, fearless one, the stars themselves will rightly testify, that my mighty, laming thunderbolt will suddenly destroy your swift horses and divine stature.” ¹⁷	—	“But if ever you transgress again, or send another substitute like him, I will show you how much hotter the thunderbolt is than your fire.” ¹⁸
—	“Phaeton is taken in by mighty Eridanus, who cleans off his blackened face.” ¹⁹	“Let his sisters bury him by the Eridanus, where he was upset. They shall weep amber tears and be changed by their grief into poplars.” ²⁰
—	“But for the horses of the Sun the weight is much too light. [...] As soon as the four-horse team observes this, they charge ahead, leaving the beaten track, not running the same path they used before.” ²¹	“And the horses soon knew I was not behind them, took the child’s measure, left the track, and wrought all this havoc.” ²²

Student Compositions in Poetry from Egypt

Surviving examples of papyri from students in Egypt provide an interesting piece of comparative evidence for understanding Sulpicius’ poem. Although these papyri postdate Sulpicius, they nonetheless reflect long-standing educational practices that had already taken root across the empire. As a result, these materials provide valuable evidence for understanding the norms of Greek literary and rhetorical education that form the context of Sulpicius’ poem. The following selection of papyri reflects a clear progression, beginning with training in mechanical skills before advancing toward rudimentary forms of literary composition.

During the later empire in Egypt, students were first taught to write rather than read extensively. Early education started with practicing letters of the Greek alphabet and developing neat handwriting. Teachers would create models, *proposita*, which students copied repeatedly to practice the script the formation of each letter. Because basic skills like being able to write one’s name were prioritized, the capability to read was considered a mark of an intermediate student who had already mastered a neat hand.²³ Beginners therefore practiced by composing *scholia minora* of the *Iliad* on papyri, where errors in alignment and handwriting were frequent.²⁴ One such example was found on a notebook of three waxed tablets, which contained a poem about Achilles. Although this story took inspiration from Homer, the focus of the exercise was to imitate the teacher’s hand rather than create an original composition.²⁵ Only after mastering skills like accuracy and fluent script did students engage in literary practices.

More advanced work demonstrates original student composition, with one papyrus fragment containing five mythological epigrams written in hexameters.²⁶ While these epigrams demonstrate some creative engagement through channeling figures from different myths, their short length lacks the complex and thorough perspective of a sustained narrative. Another papyrus fragment containing examples of poetry that intended to recreate the story of the *Odyssey*. This exercise was likely assigned by a rhetor to practice reinventing the famous topic from Homer while developing stylistic variety.²⁷ Even at this stage in education, originality was limited to brief episodes, which were constrained within established models rather than expecting students to develop personal perspectives.

The culmination of rhetorical training in Egypt is reflected by papyrus fragments practicing *êthopoia*. In one example, a student composed poetry from the perspectives of Calliope consoling Thetis and Triopas supplicating Demeter to spare his son Erysichthon, a technique parallel to Sulpicius' impersonation of Zeus.²⁸ The fluent and practiced hand suggest that the student was highly trained, indicating that this was one of the most advanced examples of rhetorical education in Egypt. A detailed comparison between the two poems can't be done here, but promises to be immensely worthwhile: what similarities and differences do we notice between the work of an Egyptian student and a student in Rome itself? The papyri also serve to highlight the rarity and value of Sulpicius' poem: Sulpicius' poem is complete and equipped with a preamble and specific historical context that allow us to better understand how and why it was composed.

Conclusion

The poem of Q. Sulpicius Maximus commemorates the longevity of the imperial education system and the Greek literary prestige it was based upon. Through close analysis of the language, subject matter, and perspective of his poem, this study has shown how Sulpicius' poem is a product of the many aspects of the imperial education system. This style of rhetorical education was deeply influenced by Homeric tradition, evident by his use of epithets, the dactylic hexameter, and the portrayal of the gods. Comparisons with contemporary interpretations of the Phaethon myth further relate to Sulpicius' unique perspective: while Ovid's Latin version adapts the mythological material, later authors like Lucian deliberately returned to writing in Greek to align with a renewed interest in classical learning. The contrast between Sulpicius' work and compositions from papyri in Greco-Roman Egypt further demonstrate how exceptional his work truly was. Sulpicius' work is a rare witness to the thrilling moment when elite Greek education left the classroom and entered the public realm through competition. It exemplifies the peak of Roman scholarly ambition and reflects how Roman education preserved Greek cultural tradition while developing a distinct imperial identity. Ultimately, Sulpicius' is an unique artifact that reveals the shared values of the Greek style of education and rhetorical excellence, while transmitting knowledge and art in the form of public display and competition for ages beyond its original context.

Endnotes

- ¹ On Q. Sulpicius Maximus and a full translation of his poem, see Brian J. Wright “The First-Century Inscription of Quintus Sulpicius Maximus: An Initial Catalog of Lexical Parallels with the New Testament.” *Bulletin for Biblical Research* 27, no. 1 (2017): 53-63. Digital file.
- ² This discussion of rhetorical education is based on ancient values expressed in Quintilian’s *Institutio Oratoria*, supplemented by modern analyses of Greek schooling. (List them here so we know what they are). References to Quintilian are taken from the selections translated in Mark Joyal, Iain McDougall, and J.C. Yardley. *Greek and Roman Education* (Publisher: 1976), p. 175.
- ³ Quintilian, *Institutio Oratoria*, 1.1.12-13.
- ⁴ Information on Atticism and renewed interest in Greek culture during the Second Sophistic can be found in Cribiore, Raffaella. *Gymnastics of the Mind*. Princeton University Press. 2001, p. 236 and Morgan, Teresa. *Literate Education in the Hellenistic and Roman Worlds*. Cambridge University Press. 1998, p. 74.
- ⁵ See Raffaella Cribiore. “Education in the Papyri.” In *The Oxford Handbook of Papyrology*, Edited by Roger S. Bagnall (Oxford) 2011, p. 331.
- ⁶ *Ibid.*, p. 332.
- ⁷ *Ibid.*, p. 333.
- ⁸ See Cribiore 2001, p. 241.
- ⁹ Quintilian, *Institutio Oratoria* cited in Joyal, Mark, Iain McDougall, and J.C. Yardley. *Greek and Roman Education*. 1976, p. 175.
- ¹⁰ Hyginus, *Fabulae*, 154
- ¹¹ Euripides, *Hippolytus*, 737-741
- ¹² Apollonius of Rhodes, *Argonautica*, IV, 595-611
- ¹³ See Georgius Knaack. *Quaestiones Phaethontaeae*. 8th ed. Edited by A. Kiessling and U.V. Wilamowitz-Moellendorff. (Berlin: Weidmannsche Buchhandlung, 1886) p. 23
- ¹⁴ Lucian, *Dialogues of the Gods*, XXV
- ¹⁵ Wright 2017, p. 58, 17-18
- ¹⁶ Ovid, *Metamorphoses*, II, 436-439
- ¹⁷ Wright 2017, p. 58, 40-43
- ¹⁸ Lucian, *Dialogues of the Gods*, XXV
- ¹⁹ Ovid, *Metamorphoses*, II, 469-471
- ²⁰ Lucian, *Dialogues of the Gods*, XXV
- ²¹ Ovid, *Metamorphoses*, II, 236-245
- ²² Lucian, *Dialogues of the Gods*, XXV
- ²³ See Raffaella Cribiore. *Writing, Teachers, and Students in Graeco-Roman Egypt*. 1996, p. 253-254
- ²⁴ See P. Oslo II 12 as cited in Cribiore 1996, p. 387
- ²⁵ See Brooklyn 37.1913E as cited in Cribiore 1996, p. 421

²⁶ See P. Heid. inv. 1271 as cited in Cribiore 1996, p. 402

²⁷ See P. Ryl. II 487 as cited in Cribiore 1996, p. 400

²⁸ See C. Graves, *Hermathena* 5 (1885) 237-57 as cited in Cribiore 1996, p. 400-401

Works Cited

- Apollonius of Rhodes. *Argonautica*. Translated by R.C. Seaton (1912). Project Gutenberg.
- Cribiore, Raffaella. "Education in the Papyri." In *The Oxford Handbook of Papyrology*, by Roger S. Bagnall. New York: Oxford University Press, 2009.
- Cribiore, Raffaella. *Gymnastics of the Mind*. Princeton: Princeton University Press, 2001.
- Cribiore, Raffaella. *Writing, Teachers, and Students in Graeco-Roman Egypt*. Atlanta: Scholars Press, 1996.
- Euripides. *Hippolytus*. Translated by Ian Johnston (2016). Richer Resources Publications. Available online: <https://johnstoniatexts.x10host.com/euripides/hippolytushtml.html>
- Fernández, Eduardo. "Retórica y Educación: una aproximación a la escuela romana" [Rhetoric and Education: An approach to the Roman school]. *Revista española de pedagogía* 80, no. 283 (2022): 475-90.
- Hyginus. *Fabulae*. Translated by Mary Grant. Theoi Texts Library. Available online: <https://www.theoi.com/Text/HyginusFabulae4.html>
- Joyal, Mark, Iain McDougall, and J.C. Yardley. *Greek and Roman Education*. London: Taylor & Francis Group, 1976.
- Knaack, Georgius. *Quaestiones Phaethontaeae*. 8th ed. Edited by A. Kiessling and U.V. Wilamowitz-Moellendorff. Berlin: Weidmannsche Buchhandlung, 1886.
- Lucian. *Dialogues of the Gods*. Translated by Henry Watson Fowler and Francis George Fowler. Oxford: Clarendon Press, 1905.
- Morgan, Teresa. *Literate Education in the Hellenistic and Roman Worlds*. Cambridge: Cambridge University Press, 1998.
- Ovid. *Metamorphoses*. Translated by Henry T. Riley, 2007. Project Gutenberg.
- Quintilian. *Institutio Oratoria*. Translated by H.E. Butler. Boston: Harvard University Press, 1920.
- Wright, Brian J. "The First-Century Inscription of Quintus Sulpicius Maximus: An Initial Catalog of Lexical Parallels with the New Testament." *Bulletin for Biblical Research* 27, no. 1 (2017): 53-63. Digital file.

TPA023B Fails to Reverse Conditioning-Related Behavioral Change in CCI Mice

By Kosei Takanashi

Abstract

Treatment of chronic pain remains a major unmet medical need, partly due to the difficulty of translating preclinical pain measures into reliable predictors of therapeutic efficacy. Preclinical assessment often relies on evoked responses, such as mechanical allodynia, as well as non-evoked conditioning-based measures; however, the extent to which these endpoints reflect overlapping or distinct biological and behavioral mechanisms is unclear. In this study, we examined relationships among previously published baseline sensitivity, post-treatment evoked responses, and conditioning-based outcomes in a chronic constriction injury (CCI) mouse model following administration of TPA023B, a subtype-selective GABA_A receptor partial agonist. We found strong positive correlations across post-treatment time points, suggesting high temporal stability in von Frey responses. Baseline mechanical sensitivity moderately predicted post-treatment outcomes but showed minimal association with conditioning-derived measures, indicating partial independence between evoked and non-evoked domains. Conditioning responses were highly consistent within individuals across sessions and treatments, reflecting stable behavioral phenotypes. To explore predictive capacity, supervised machine learning models—including a neural network, ridge regression, and lasso regression—were trained and evaluated. While the neural network achieved the lowest mean absolute error, all models yielded negative R^2 values. These findings underscore the redundancy of dense time-course measurements, the importance of baseline sensitivity, and the challenges of applying machine learning to small, variable preclinical datasets.

Introduction

Effective treatment for chronic pain represents a significant unmet medical need and affects hundreds of millions of individuals worldwide. Despite extensive investment in analgesic drug development, the clinical pipeline for novel pain therapeutics has seen limited success in recent decades (Dworkin et al., 2007; Finnerup et al., 2010). Many existing treatments, including opioids, are associated with serious adverse effects, such as tolerance, dependence, and risk of misuse, while non-opioid alternatives often provide inadequate relief (Finnerup et al., 2010; Woolf, 2010). A major challenge in developing safer pain medications is the biological complexity and heterogeneity of pain, which involves sensory, emotional, and motivational components. Furthermore, high failure rates in clinical trials suggest that preclinical models often fail to predict human efficacy. This translational gap has slowed advancement in pain therapeutics. Improving the predictive validity of preclinical pain assessment is therefore a critical need for the field.

In preclinical research, pain is most commonly assessed using behavioral assays that measure responses to controlled stimuli. Evoked pain measures, such as mechanical allodynia assessed with von Frey filaments and thermal sensitivity measured using the Hargreaves assay,

quantify reflexive responses to external stimuli (Chaplan et al., 1994). These methods are widely used due to their relative simplicity, reproducibility, and sensitivity to pharmacological intervention. However, evoked measures primarily capture stimulus-evoked hypersensitivity and may not reflect ongoing or spontaneous pain states (Mogil, 2009). To address this limitation, non-evoked and conditioning-based assays have been developed to assess pain-related behaviors in the absence of direct stimulation (King et al., 2009; Navratilova et al., 2012). These approaches aim to capture affective and motivational dimensions of pain but often exhibit greater variability (Tappe-Theodor et al., 2019). As a result, evoked and non-evoked measures are frequently collected together, despite uncertainty regarding how they relate to one another.

Bridging the gap between evoked and non-evoked pain measures remains an unresolved challenge in pain research. Prior studies have shown that outcomes from different assays do not always align, suggesting that distinct biological and neural mechanisms may underlie these measures (Mogil, 2009; Tappe-Theodor et al., 2019). While some overlap likely exists, the degree to which evoked hypersensitivity predicts non-evoked or conditioning-based responses is unclear (King et al., 2009). This raises important questions about redundancy among commonly collected endpoints and their relative value for predicting drug efficacy (Mogil, 2009; Woolf, 2010). This raises the question of whether these tests overlap or measure completely different things to better predict if a drug will have efficacy in humans. To solve this, data science and machine learning approaches offer a powerful framework for systematically evaluating relationships among complex behavioral datasets.

In this study, we integrated correlation analysis and supervised machine learning to systematically evaluate a longitudinal dataset from a chronic constriction injury (CCI) mouse model treated with TPA023B. By analyzing the relationships among baseline mechanical sensitivity, post-treatment evoked time courses, and conditioning-based outcomes, we sought to address the translational gap and predictive limitations inherent in current preclinical pain assessment. We hypothesized that applying these computational approaches would clarify the biological relationships between distinct behavioral domains, reveal redundancies in dense time-course measurements, and inform strategies to improve translational prediction. This work addresses critical gaps in the literature by evaluating whether commonly collected evoked measures can reliably predict the affective-motivational dimensions of pain, thereby identifying which behavioral endpoints provide the most independent and translatable information for analgesic drug development.

Materials and Methods

Dataset Description and Preprocessing

The dataset analyzed in this study (n=12) was derived from a previously published preclinical pain experiment using a chronic constriction injury (CCI) model (Neumann et al., 2021). It is a well-established model of neuropathic pain characterized by persistent mechanical hypersensitivity with behavioral outcomes quantified using evoked and conditioning-based

assays (Austin et al., 2012; Woolf, 2010). The primary goals of the initial data collection were to characterize the longitudinal effects of the alpha 2/alpha 3 GABA{A} receptor modulator TPA023B on sensory hypersensitivity and to evaluate its efficacy in alleviating the tonic affective component of chronic pain. Specifically, mechanical allodynia measurements were collected at baseline and across multiple post-dose time points following administration of TPA023B (0.3 mg/kg), administered either per oral (p.o.) via gavage needle for systemic assessment or intrathecally (i.t.) via spinal injection between the L5 and L6 vertebrae to target spinal GABA{A} receptors, on relationships among previously published baseline sensitivity, post-treatment evoked responses, and conditioning-based outcomes in a chronic constriction injury (CCI) mouse model, generating a longitudinal evoked pain time course (Neumann et al., 2021). In parallel, conditioning-based measures were obtained under both vehicle and drug conditions, including pre-conditioning and post-conditioning response values (Neumann et al., 2021).

To facilitate comparative and predictive analyses, derived change variables were computed as the difference between pre- and post-conditioning values for each treatment condition (Pre–Post-Veh and Pre–Post-Drug). All variables were organized into a structured tabular format, and only numeric features relevant to behavioral outcomes were retained. These features included baseline mechanical sensitivity and a five-point time course of post-treatment reflexive responses. They also encompassed raw chamber occupancy times from pre-conditioning and post-conditioning sessions for both vehicle and TPA023B. Finally, the set included the calculated change scores that measure natural behavioral shifts versus drug-induced relief. No imputation was performed beyond exclusion of incomplete entries, ensuring that analyses were conducted on directly observed measurements. This preprocessing approach enabled consistent feature representation across correlation and machine learning analyses.

Machine Learning Analysis

To explore the predictive capacity of the behavioral dataset, three supervised machine learning models were implemented: a feedforward neural network, used to detect complex nonlinear patterns through interconnected nodes; ridge regression utilizing L2 regularization to prevent over-reliance on individual data points; and lasso regression, employing L1 regularization to simplify the model through feature selection. These models were selected to represent both nonlinear (neural network) and linear regularized approaches, allowing comparison of model complexity and interpretability. Ridge regression applies L2 regularization-the penalty is based on the square of the coefficients. This keeps the weights small and spread out, ensuring no single variable dominates the prediction and reduces overfitting by penalizing large coefficients. The penalty is based on the absolute value of the coefficients. This is a stricter penalty that can force less useful features to zero, effectively "firing" them from the model to keep it simple- to promote sparsity and implicit feature selection.

Model performance was evaluated using standard regression metrics, including mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and the

coefficient of determination (R^2). MAE quantifies average absolute prediction error, MSE and RMSE emphasize larger errors, and R^2 measures the proportion of variance explained relative to a mean-based baseline. These metrics provide complementary insight into both predictive accuracy and explanatory power across modeling approaches.

Table 1

Model	Approach	How it works	Main Benefit
Neural Network	non-linear	Uses interconnected "nodes" to find hidden, messy patterns.	Best at catching complex, non-linear relationships.
Ridge Regression	linear	Shrinks the influence of all data points so no single one takes over.	Prevents the model from relying too much on one piece of info.
Lasso Regression	linear	Acts like a filter that completely removes the least helpful data.	Simplifies the model by "firing" useless features.

To assess relationships between key experimental variables, a correlation matrix was computed and visualized as a heatmap (Fig.2). Analyses were performed using Python (pandas and seaborn libraries). The correlation matrix was visualized using `seaborn.heatmap()` with annotation of correlation coefficients (formatted to two decimal places), and a cool-warm diverging colormap to represent the strength and direction of correlations. This visualization enabled rapid identification of strongly positive and negative relationships among experimental measures.

A subset of relevant numerical variables was selected from the full dataset and stored in a DataFrame (df_relevant). The following columns included the initial physical sensitivity (Baseline-MA) and a time course of sensitivity responses measured at one hour (1hpost0.3mg/kgTPA023Bit), ninety minutes (1.5hpost0.3mg/kgTPA023Bit), two hours (2hpost0.3mg/kgTPA023Bit), three hours (3hpost0.3mg/kgTPA023Bit), and four hours (4hpost0.3mg/kgTPA023Bit) after drug administration. The dataset also included the starting chamber times for the control group (Pre-conditioning-Vehicle(s)) and the drug group (Pre-conditioning-TPA023B(s)), along with the final times for those same areas after training (Post-conditioning-Vehicle(s) and Post-conditioning-TPA023B(s)). Finally, we included the calculated scores for natural behavior shifts (Pre-Post-Veh) and actual drug-induced relief (Pre-Post-Drug). Pairwise Pearson correlation coefficients were calculated using the `corr()` function in the pandas library, generating a correlation matrix that quantifies linear relationships between all variable pairs.

Results

Data Preprocessing (normalized CPP)

TPA023B produced a robust reversal of conditioning-related behavioral change in CCI mice, shifting responses from a large negative pre–post change under vehicle to a strong positive change with drug treatment (Fig. 1). In contrast, naïve mice showed only modest changes, with minimal separation between vehicle and TPA023B. These results indicate that TPA023B preferentially modulates pain-related behavior in the injured (CCI) state, with limited effects in naïve animals (Fig. 1).

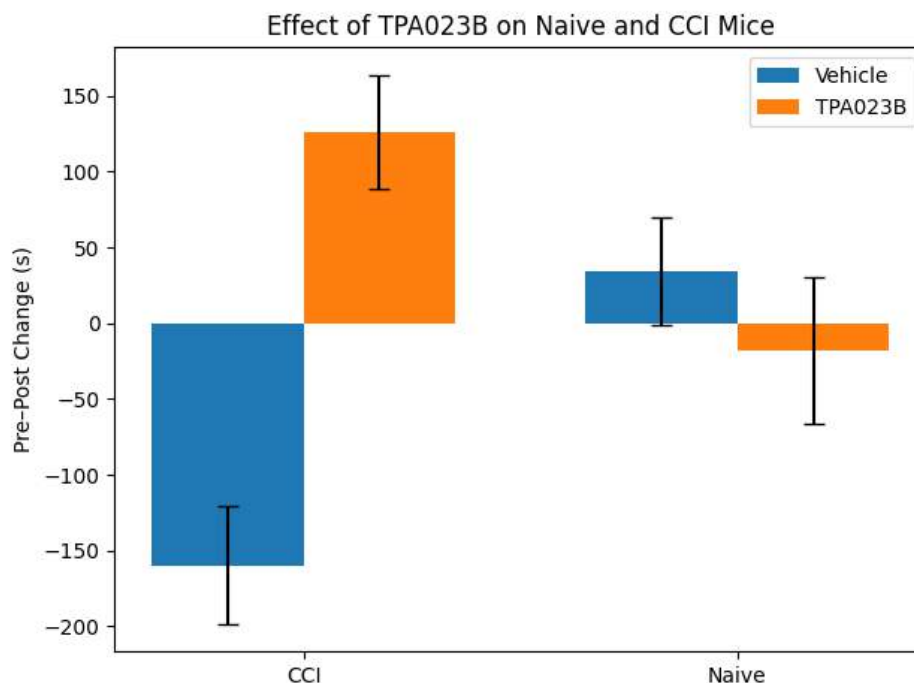


Figure 1: Effect of TPA023B on naïve and CCI mice.

Acronyms

TPA023B — GABAA_A receptor–modulating compound tested for analgesic and conditioning effects. CCI — Chronic Constriction Injury (neuropathic pain model). Pre–Post — Difference between post-conditioning and pre-conditioning time spent in a chamber. s — seconds. Pre–post change in time spent (seconds) in the drug- or vehicle-paired chamber is shown for naïve mice ($n = 12$) and mice subjected to chronic constriction injury (CCI, $n = 9$). TPA023B produced a strong positive conditioning effect in CCI mice compared with vehicles, whereas naïve mice showed minimal or opposite changes, suggesting that the rewarding or relief-associated effects of TPA023B are more pronounced under neuropathic pain conditions.

Strong Positive Correlations Among Post-Treatment Time Points

Strong positive correlations were observed among the post-treatment time-series variables (1 h, 1.5 h, 2 h, 3 h, and 4 h post-TPA023B) (Fig.2). Adjacent and non-adjacent time points showed consistently high correlation coefficients, indicating that animals with higher responses at one time point tended to maintain higher responses at subsequent time points. For example, mechanical sensitivity at the 3-hour mark showed a near-perfect correlation with the 4-hour mark ($r = 0.94$) and a very strong relationship with the 2-hour mark ($r = 0.70$). Even non-adjacent intervals, such as 1 hour versus 4 hours post-dose, maintained a high correlation ($r = 0.69$). These specific values suggest a stable and continuous temporal treatment effect throughout the observation window.

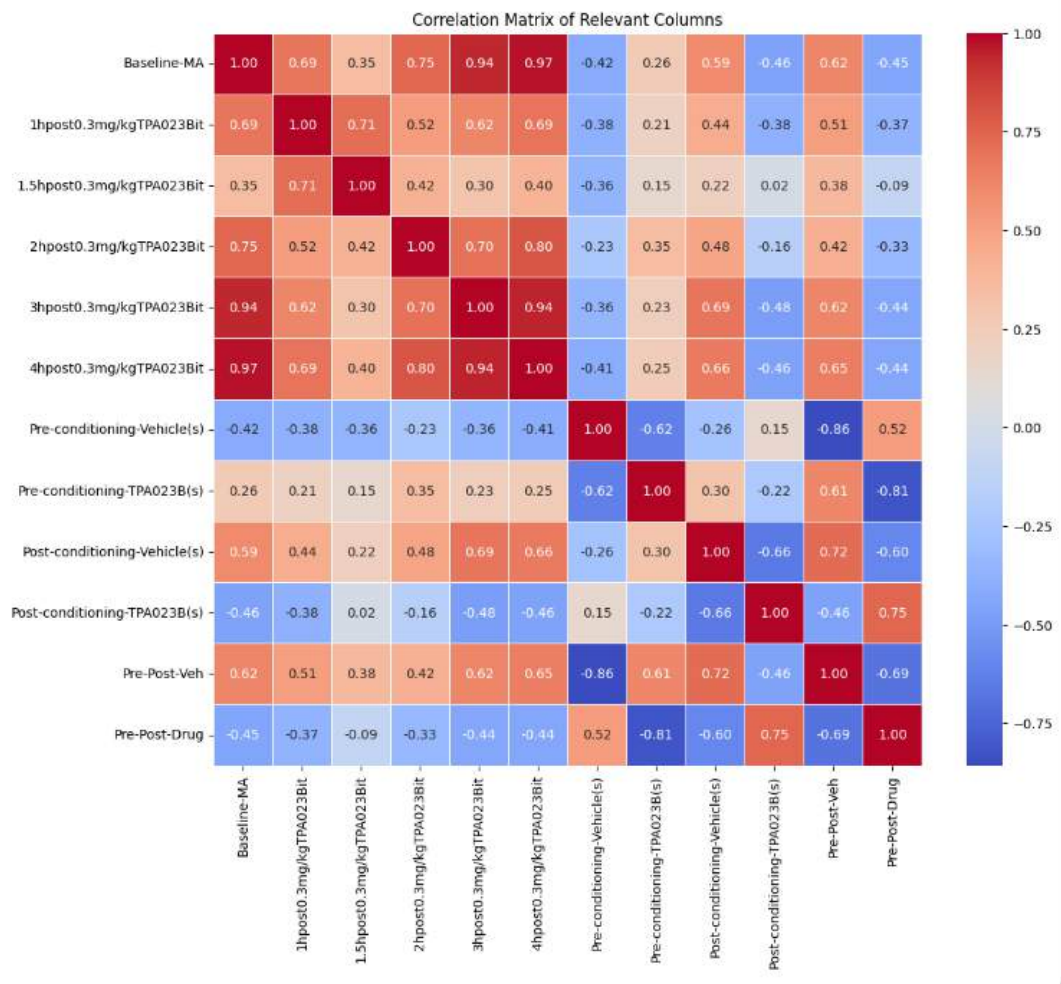


Figure 2 :Correlation Matrix of Behavioral Responses and Physiological Baselines Following TPA023B Administration.

Acronyms

MA: Mechanical Allodynia (pain from non-painful stimuli), ML: Machine Learning (AI that learns from data), CCI: Chronic Constriction Injury (a pain model), TPA023B: A specific compound or drug being studied.

Overall Heatmap Figure Description

The heatmap visually shows how strongly different experimental measurements are related to each other. Colors indicate if two measurements tend to change together (positive correlation) or in opposite ways (negative correlation). This is an effective visualization tool for establishing and exploring patterns and relationships within a data/feature set.

Correlations with Baseline Measurements

Baseline mechanical allodynia (Baseline-MA) showed moderate to strong positive correlations with post-treatment measurements (approximately $r = 0.6$ – 0.7), indicating that baseline sensitivity partially predicts post-treatment response magnitude (Fig.2). In contrast, Baseline-MA showed weak or negligible correlations with conditioning-related variables (pre-conditioning, post-conditioning, and pre–post difference measures), suggesting that baseline sensitivity is largely independent of conditioning responses (Fig. 2).

To assess the relationships across the dataset, a correlation matrix was generated and visualized as a heatmap (Fig. 2). To interpret this figure, the reader should observe the x and y axes, which both list the categorical variables of the study; each intersecting square displays a correlation coefficient, where darker, more saturated colors indicate stronger relationships.

Pre- and Post-Conditioning Consistency

Strong positive correlations were observed between pre- and post-conditioning measures within both treatment groups (Fig.2): Pre-conditioning-Vehicle vs. Post-conditioning-Vehicle ($r \approx 0.72$), and Pre-conditioning-TPA023B vs. Post-conditioning-TPA023B ($r \approx 0.70$). These findings indicate that individuals tend to maintain relative response levels across conditioning sessions, reflecting stable behavioral phenotypes.

Behavior of Derived Pre–Post Difference Variables

As expected from their mathematical definition, the difference variables showed characteristic correlation patterns (Fig.2): Pre-Post-Veh- a baseline to measure any natural shifts in behavior over time that are not caused by the actual drug treatment- was strongly negatively correlated with Pre-conditioning-Vehicle ($r \approx -0.81$) and moderately positively correlated with Post-conditioning-Vehicle ($r \approx 0.32$). Pre-Post-Drug was strongly negatively correlated with Pre-conditioning-TPA023B ($r \approx -0.84$) and moderately positively correlated with Post-conditioning-TPA023B ($r \approx 0.35$). These relationships suggest that change scores are driven by both initial and final values, and that higher baseline responses tend to yield smaller net increases (or larger decreases).

Inter-Treatment Relationships

Moderate positive correlations were observed between vehicle and drug conditions (Fig.2), Pre-conditioning-Vehicle vs. Pre-conditioning-TPA023B ($r \approx 0.44$), Post-conditioning-Vehicle vs. Post-conditioning-TPA023B ($r \approx 0.45$). Additionally, Pre-Post-Veh

and Pre-Post-Drug were moderately correlated ($r \approx 0.46$), suggesting that individuals who exhibit larger changes under vehicle conditioning also tend to show larger changes under drug conditioning. This indicates shared individual sensitivity or common underlying behavioral mechanisms across conditions.

Summary of Heatmap Findings

Overall, the heatmap reveals: high temporal stability in post-treatment responses, partial predictive value of baseline sensitivity for drug response, consistency of conditioning responses within individuals, and shared individual variability across vehicle and drug treatments. These relationships highlight both redundant variables and biologically meaningful patterns that are important for feature selection and predictive modeling.

Strategy for Exploring Additional Machine Learning Analysis and Model Performance

The results from running a neural network, ridge regression, and lasso regression are shown in Table 1. The neural network achieved the lowest overall error, with an MAE of 108.46, MSE of 16,792.30, and RMSE of 129.59, while ridge regression produced an MAE of 113.49, MSE of 18,241.19, and RMSE of 135.06. Lasso regression showed the highest error, with an MAE of 117.07, MSE of 20,054.68, and RMSE of 141.61. Despite these differences, all three models yielded negative R^2 values (-0.17 for the neural network, -0.27 for ridge regression, and -0.39 for lasso regression), indicating that none outperformed a baseline model that predicts the mean of the target variable.

Table 2

<u>ML Method</u>	<u>Mean absolute error (MAE)</u>	<u>Mean squared error (MSE)</u>	<u>Root mean squared error (RMSE)</u>	<u>Coefficient of determination (R^2)</u>
Neural Network	108.46	16792.30	129.59	-0.17
Ridge Regression	113.49	18241.19	135.06	-0.27
Lasso Regression	117.07	20054.68	141.61	-0.39

Table 2: Performance of supervised machine learning models predicting behavioral outcomes

Summary of predictive performance for supervised machine learning models trained on evoked and conditioning-based behavioral features. Models included a feedforward neural network, ridge regression, and lasso regression. Specifically, these models were trained to predict the pre-post change in chamber occupancy (affective pain relief) using reflexive mechanical sensitivity data as input features. Model validation was performed using a stratified 80/20 train-test split; this approach ensured that the ratio of naive ($n = 12$) to CCI ($n = 9$) animals remained consistent across both the training and testing subsets to maintain group representation.

Performance was evaluated using mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R^2). Although the neural network achieved the lowest overall error metrics, all models yielded negative R^2 values, indicating that none outperformed a baseline predictor based on the mean outcome. These findings highlight the limited predictive capacity of the current feature set and underscore challenges associated with small, variable preclinical pain datasets.

Discussion

The correlation data highlight an interesting pattern of temporal redundancy, despite the traditional focus on detailed von Frey time courses (1h, 1.5h, 2h, 3h, and 4h post-dose), measurements across different post-dose time points were highly correlated ($r \sim 0.94$ between late-stage points). This suggests that once a drug effect is established, subsequent measurements provide diminishing returns of independent information, as the stability of the mechanical sensitivity across hours indicates that the reflexive response is relatively static during the testing window (He et al., 2012) (Fig.2). This raises a practical question about the value of collecting dense time-course data if each measurement provides limited independent information. Baseline mechanical sensitivity, in contrast, consistently predicted post-dose responses, indicating that a single pre-injury or pre-treatment measure may carry significant predictive power (Austin et al., 2012) (Fig. 2). For the field, this suggests that experimental designs could potentially be optimized to reduce redundancy, focusing on key predictive time points while expanding measurement domains, such as including complementary modalities like heat sensitivity assays (e.g., Hargreaves), to capture distinct aspects of nociceptive processing. Exploring these additional phenotypic dimensions could improve understanding of pain mechanisms and inform more efficient behavioral workflows.

From a computational perspective, the failure of the machine learning models to achieve a positive R^2 serves as a critical diagnostic of current preclinical data structures. The negative values—ranging from -0.17 to -0.39—confirm that the models were unable to find a generalizable "rule" connecting reflexive sensitivity to affective relief. This poor performance is likely a byproduct of high biological "noise" outweighing the "signal" within a small sample size ($n=12$). In behavioral studies, individual differences often create a high variance that causes complex models like neural networks to overfit, essentially "memorizing" the quirks of the specific training group rather than learning a general biological principle (Lötsch & Ultsch, 2018).

This suggests that the machine learning approach is not "failing" in the traditional sense; rather, it is identifying a significant "feature gap." The current set of reflexive features—while standard in the field—does not appear to contain enough unique information to reliably predict how an animal will behave in a complex preference task. To bridge this gap, future models must move beyond simple thresholds and incorporate richer, multidimensional datasets. Integrating high-resolution video tracking or longitudinal physiological monitoring could provide the "deep" input necessary for neural networks to function effectively (Winslow et al., 2022). Ultimately,

shifting the focus from redundant longitudinal time points toward greater feature diversity will be essential for developing the predictive models needed to accelerate the translation of novel analgesics.

Conclusion

In this study, we examined both evoked and non-evoked pain behaviors in a preclinical mouse model of chronic constriction injury (CCI), assessing responses under vehicle and TPA023B treatment conditions (n=12). We analyzed previously collected measurements of time-course mechanical allodynia using von Frey filaments and conditioning-based behavioral assays to capture non-evoked pain (Assareh et al., 2025; He et al., 2012; Neumann et al., 2021). Correlation analyses revealed strong relationships across von Frey time points, moderate predictive value of baseline scores for post-dose responses, and consistency between pre- and post-conditioning measures (Fig. 2). Derived change variables behaved as expected mathematically, and moderate correlations between vehicle and drug conditions suggest shared individual sensitivity (Fig. 2). Machine learning approaches, including neural networks, Ridge regression, and Lasso regression, were applied to predict behavioral outcomes (Langford et al., 2025; Lötsch & Utsch, 2018). While the neural network achieved the lowest mean absolute error, overall performance remained poor, and R^2 values were negative across all models (Table 1). These results collectively indicate that current features and sample sizes are insufficient to fully model the underlying biological mechanisms driving pain behavior.

Works Cited

- Assareh, N., Sokolaj, E. E., Sadia, S., Anderson, K. E., Frith, C., Walls, O. B., Mitchell, V. A., Vaughan, C. W., & Winters, B. L. (2025). Fear-learning is altered in a mouse neuropathic pain model. *Frontiers in Pain Research*, 6, 1648374. <https://doi.org/10.3389/fpain.2025.1648374>
- Austin, P. J., Wu, A., & Moalem-Taylor, G. (2012). Chronic constriction of the sciatic nerve and pain hypersensitivity testing in rats. *Journal of Visualized Experiments: JoVE*, (61), 3393. <https://doi.org/10.3791/3393>
- Bennett, G. J., & Xie, Y.-K. (1988). A peripheral mononeuropathy in rats that produces disorders of pain sensation like those seen in man. *Pain*, 33(1), 87–107. [https://doi.org/10.1016/0304-3959\(88\)90209-6](https://doi.org/10.1016/0304-3959(88)90209-6)
- Breivik, H., Collett, B., Ventafridda, V., Cohen, R., & Gallacher, D. (2006). Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment. *European Journal of Pain*, 10(4), 287–333. <https://doi.org/10.1016/j.ejpain.2005.06.009>
- Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, 3(3), 223–230. <https://doi.org/10.1016/j.bpsc.2017.11.007>
- Chaplan, S. R., Bach, F. W., Pogrel, J. W., Chung, J. M., & Yaksh, T. L. (1994). Quantitative assessment of tactile allodynia in the rat paw. *Journal of Neuroscience Methods*, 53(1), 55–63. [https://doi.org/10.1016/0165-0270\(94\)90144-9](https://doi.org/10.1016/0165-0270(94)90144-9)
- Dworkin, R. H., O'Connor, A. B., Backonja, M., Farrar, J. T., Finnerup, N. B., Jensen, T. S., Kalso, E. A., Loeser, J. D., Miaskowski, C., Nurmikko, T. J., Portenoy, R. K., Rice, A. S. C., Stacey, B. R., Treede, R.-D., Turk, D. C., & Wallace, M. S. (2007). Pharmacologic management of neuropathic pain: Evidence-based recommendations. *Pain*, 132(3), 237–251. <https://doi.org/10.1016/j.pain.2007.08.033>
- Finnerup, N. B., Sindrup, S. H., & Jensen, T. S. (2010). The evidence for pharmacological treatment of neuropathic pain. *Pain*, 150(3), 573–581. <https://doi.org/10.1016/j.pain.2010.06.019>
- He, Y., Tian, X., Hu, X., Porreca, F., & Wang, Z. J. (2012). Negative reinforcement reveals non-evoked ongoing pain in mice with tissue or nerve injury. *The Journal of Pain*, 13(6), 598–607. <https://doi.org/10.1016/j.jpain.2012.03.011>
- King, T., Vera-Portocarrero, L., Gutierrez, T., Vanderah, T. W., Dussor, G., Lai, J., Fields, H. L., & Porreca, F. (2009). Unmasking the tonic-aversive state in neuropathic pain. *Nature Neuroscience*, 12(11), 1364–1366. <https://doi.org/10.1038/nn.2407>
- Langford, D. J., Reichel, J. F., Zhong, H., Basseri, B. H., Koch, M. P., Kolady, R., Liu, J., Sideris, A., Dworkin, R. H., Poeran, J., & Wu, C. L. (2025). Machine learning research methods to predict postoperative pain and opioid use: A narrative review. *Regional Anesthesia and Pain Medicine*, 50(2), 102–109. <https://doi.org/10.1136/rapm-2024-105603>

- Lötsch, J., & Utsch, A. (2018). Machine learning in pain research. *Pain*, 159(4), 623–630. <https://doi.org/10.1097/j.pain.0000000000001118>
- Mogil, J. S. (2009). Animal models of pain: Progress and challenges. *Nature Reviews. Neuroscience*, 10(4), 283–294. <https://doi.org/10.1038/nrn2606>
- Navratilova, E., Xie, J. Y., Okun, A., Qu, C., Eyde, N., Ci, S., Ossipov, M. H., King, T., Fields, H. L., & Porreca, F. (2012). Pain relief produces negative reinforcement through activation of mesolimbic reward-valuation circuitry. *Proceedings of the National Academy of Sciences of the United States of America*, 109(50), 20709–20713. <https://doi.org/10.1073/pnas.1214605109>
- Neumann, E., Küpfer, L., & Zeilhofer, H. U. (2021). The $\alpha 2/\alpha 3$ GABAA receptor modulator TPA023B alleviates not only the sensory but also the tonic affective component of chronic pain in mice. *Pain*, 162(2), 421–431. <https://doi.org/10.1097/j.pain.0000000000002030>
- Tappe-Theodor, A., King, T., & Morgan, M. M. (2019). Pros and Cons of Clinically Relevant Methods to Assess Pain in Rodents. *Neuroscience and Biobehavioral Reviews*, 100, 335–343. <https://doi.org/10.1016/j.neubiorev.2019.03.009>
- Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- Vierck, C. J., Hansson, P. T., & Yezierski, R. P. (2008). Clinical and pre-clinical pain assessment: Are we measuring the same thing? *Pain*, 135(1–2), 7–10. <https://doi.org/10.1016/j.pain.2007.12.008>
- Winslow, B. D., Kwasinski, R., Whirlow, K., Mills, E., Hullfish, J., & Carroll, M. (2022). Automatic detection of pain using machine learning. *Frontiers in Pain Research*, 3, 1044518. <https://doi.org/10.3389/fpain.2022.1044518>
- Woolf, C. J. (2010). Overcoming obstacles to developing new analgesics. *Nature Medicine*, 16(11), 1241–1247. <https://doi.org/10.1038/nm.2230>

Gene Therapies and Their Influence on Inherited Retinal Disorders By Tammy Berinstein

1.0 Abstract

Inherited retinal disorders (IRDs) affect the retina and greatly impact daily life. IRDs are genetic disorders, meaning they are unable to be cured through conventional treatments and medicines. To expand the available treatments for IRDs and other genetic disorders, new gene therapy technologies are being researched and created. This review analyzes the limited treatment options for IRDs and the trials being conducted to combat this issue. Currently, LUXTRNA is the only approved gene therapy for IRDs, and it is a gene replacement therapy instead of a gene editing therapy, meaning it adds a healthy copy of the gene without fixing the mutated one. Pre-clinical trials are being conducted for CRISPR Cas9 nucleases and CRISPR base editing. CRISPR base editing is more promising than CRISPR Cas9 nucleases, as CRISPR Cas9 nucleases have a higher risk for off-target edits. Additionally, gene editing can raise safety concerns; however, if these technologies advance and are further studied, they have the potential to revolutionize the medical industry and greatly help those who live with IRDs.

Key Words

Gene editing, gene therapies, inherited retinal disorders (IRDs), gene replacement, CRISPR base editing, CRISPR Cas9 nucleases

2.0 Introduction

Around 350 million people worldwide, which is approximately 4.3% of all people, live with a rare disorder, with 80% of them being due to genetics (*Rare Genetic Diseases*). Around 95% of these disorders have no known treatment and can have significant effects on daily life. While more traditional evaluation methods are often used, genetic testing is the most concrete way to diagnose a genetic disorder. The best way to treat genetic disorders is to use gene therapies, particularly gene editing. Research to develop different types of gene editing, like CRISPR Cas9 nucleases and CRISPR base editing, is being conducted for many different genetic disorders. Furthermore, inherited retinal disorders are one of the most common genetic disorders, and they affect the eyes. More than 250 genes are known to cause IRDs.

Overall, there are very limited treatment options for genetic disorders. There is only one approved CRISPR Cas9 nucleases treatment, called CASGEVY (*Patient Information | CASGEVY® (Exagamglogene Autotemcel)*). CASGEVY must be performed ex vivo and does not fix the adult hemoglobin; it solely allows for the production of healthy fetal hemoglobin, emphasizing its limitations. For IRDs, there is only one approved gene therapy, called LUXTRNA (“LUXTRNA® (Voretigene Neparvovec-Rzyl)”). LUXTRNA is a gene replacement therapy, meaning it does not fix the mutated copy; it uses a viral vector to add a healthy copy instead. It repairs vision but does not fix the root of the problem. To fix this issue, gene editing trials are being performed, and many are for IRDs.

This review aims to analyze the current treatment options for IRDs, and it describes studies being conducted to create new and improved treatment options.

3.0 Methodology

This literature review used Google and Google Scholar to find the sources. Papers analyzed were in English. The papers had a focus on IRDs, gene therapies, genetic testing, or conventional treatments for genetic disorders. Key search terms were “IRDs,” “Inherited Retinal Disorders,” “Gene therapies for IRDs,” “CRISPR Cas9 nucleases,” “CRISPR base editing,” and “Gene replacement.”

4.0 IRDs and Their Negative Effects on People

DNA contains instructions for how to make proteins, and with a mutation, an issue can arise in their production (Alberts B et al.). These problems can cause different conditions, which typically have limited treatment options. Currently, many ways to treat genetic disorders are being researched, but as of August 13, 2024, only 12 are approved, combating only 11 genetic disorders in Europe and/or the USA (Hwu). Many genetic disorders affect the eyes, but out of the 12 regulated gene therapies, only one treats IRDs.

4.1 Definition, Known Causes, and Prevalence

IRDs are a class of genetic disorders that affect the retinal cells, leading to severe vision loss or blindness (“Inherited Retinal Diseases”). Several gene mutations cause different types of IRDs, which vary in onset of disease, symptoms, and progression rates (*Inherited Retinal Disorders (IRD) | Boston Children’s Hospital*). Many IRDs are autosomal recessive disorders, which are contracted when two copies of a recessive gene are inherited. Furthermore, they are caused by a Mendelian mutation in around 1 of at least 300 genes; the exact number of genes that cause IRDs is unknown. IRDs are a common type of genetic disorder, as in 2020, around 5.5 million people had an IRD (Hanany et al.).

4.2 Different Types of IRDs and Their Varying Effects

A relatively common form of IRD is Leber congenital amaurosis (LCA), which occurs in 2 to 3 newborns out of 100,000 (*Leber Congenital Amaurosis; Inherited Retinal Disorders (IRD) | Boston Children’s Hospital*). It is present since birth, causing severe vision impairment, which has the potential to worsen over time. Additionally, LCA is associated with photophobia (an increased sensitivity to light), nystagmus (involuntary eye movements), and hyperopia (extreme farsightedness). Instead of expanding and contracting within the eyes, the pupils may not respond, or if they do, they respond more slowly than a normal eye. Repeated poking, pressing, and rubbing of the eyes with a knuckle or a finger, which is common with LCA, causes phosphenes, flashes of light. This is called Franceschetti's oculo-digital sign. LCA results from mutations in at least 20 essential genes; however, in about 30% of cases, the cause is still unknown. In addition, it usually follows an autosomal recessive pattern.

Another common class of IRDs is retinitis pigmentosa (RP) (*Inherited Retinal Disorders (IRD) | Boston Children's Hospital; Retinitis Pigmentosa | National Eye Institute*). RP causes retinal cells to break down over time, making it a degenerative disease that eventually leads to vision loss. The first symptom typically noticed with RP is a loss of vision at night. Later, it reduces peripheral vision until the person can only see through their central vision tunnel. RP can also bring sensitivity to bright light and color blindness. In addition, RP increases the risk of other eye-related problems, like cataracts and refractive errors.

Stargardt disease is the most common form of inherited macular degeneration; it affects around 30,000 people in the United States. It is usually noticed during the adolescent stage. With Stargardt disease, photoreceptors in the macula (a part of the retina that allows for central vision) are lost, leading to a loss of central vision. In addition, the retinal pigment epithelium is damaged. In the retinal pigment epithelium, there are yellow flecks, which are pockets of lipofuscin, a normal pigment that accumulates in the retina; however, with Stargardt disease, the amount of lipofuscin is excessive. In many cases, Stargardt disease is caused by mutations in the *ABCA4* gene. Less commonly found, it can be caused by a mutation in the *ELOVL4* gene.

Although it is rare to have an IRD, for those who do, it can greatly impact their daily life (*Inherited Retinal Disorders (IRD) | Boston Children's Hospital; Road et al.*). Furthermore, current treatment options are limited, with many only addressing the symptoms. Given that there is only one approved gene therapy for IRDs (“LUXTURN[®] (Voretigene Neparvovec-Rzyl)”), it is crucial to keep expanding options for more types of IRDs.

4.3 Limited Diagnosing and Treatment Options

It is difficult to diagnose IRDs through examining phenotypes because there are many different variations of symptoms for each one, and different IRDs may appear to have similar phenotypes even though something different is occurring internally (Tatour and Ben-Yosef). Conventional treatment options focus on treating the symptoms of the IRD; however, these treatments are unable to cure the IRD as they do not fix the problem itself, the gene mutation. For IRDs, many conventional options, like glasses, are not very helpful as glasses bend light; they do not fix phototransduction. Some options, like tinted glasses, may be useful to help people with photophobia; however, photophobia is only one symptom of many caused by an IRD.

5.0 Gene Replacement for IRDs: Its Efficacy Over Traditional Options

To treat the core of the problem, the gene mutation, gene replacement techniques could be a potential therapeutic option. Gene replacement provides a healthy copy of the mutated gene to the cell via a viral vector (Battu et al.). Gene replacement therapies are promising treatment options as they enable healthy proteins to be produced from a functional gene. For example, if there is a loss-of-function mutation, a viral vector carrying a healthy form of the gene of interest would be supplied. When a loss-of-function mutation occurs, the gene affected is not encoded properly, causing the cells to produce a non-functional protein. After a viral vector encoding a corrected copy of the gene of interest is delivered, a healthy protein will be generated from this

supplemented gene. An example of an approved gene replacement therapy is LUXTURNA, which enables the treatment of a subset of IRDs (“LUXTURNA® (Voretigene Neparvovec-Rzyl)”).

5.1 LUXTURNA, the First Approved Gene Therapy for IRDs

LUXTURNA is the first FDA-approved gene therapy to help people with IRDs, particularly for those with mutations in both copies of their *RPE65* gene (“Inherited Retinal Diseases”; “LUXTURNA® (Voretigene Neparvovec-Rzyl)”). LUXTURNA is a one-time gene therapy that helps restore vision.

The *RPE65* gene provides instructions for making an important protein for vision, located in the epithelium layer of the eye (*RPE65 Gene*). The protein encoded by the *RPE65* gene plays a key role in phototransduction; it converts the molecule from 11-cis-retinal to all-trans-retinal. A mutation in the *RPE65* gene causes different genetic diseases, like LCA, fundus albipunctatus, and RP.

LUXTURNA is injected subretinally (“LUXTURNA® (Voretigene Neparvovec-Rzyl)”). One eye is treated before the other, with at least 6 days in between the treatments. The recommended dosage is 1.5×10^{11} vector genomes, administered in a total volume of 0.3mL. It is recommended to take corticosteroids, an anti-inflammatory medicine, for 3 days before treatment and 4 days after. The recommended dosage is 1mg per kilogram of mass per day, with a maximum of 40mg per day. After the next 10 days, it is recommended to take a smaller dose. This treatment is not recommended for people below 12 months old; however, between the ages of 4 and 17, there was no significant difference in results. Additionally, the vector has not yet been tested on pregnant women.

LUXTURNA is administered in a surgical setting by a surgeon who is accustomed to performing intraocular surgeries (“LUXTURNA® (Voretigene Neparvovec-Rzyl)”). To start, the eyes are dilated, and the patient is given anesthesia. Afterward, a topical microbiocide is applied to the eyelid, cornea, and conjunctiva to prevent infection. A vector containing a healthy copy of the *RPE65* gene is injected into retinal cells using a syringe. Even after the procedure, vector DNA from the gene replacement therapy is often still present in tears for a limited time.

LUXTURNA’s success rate was measured using a multi-luminance mobility testing (MLMT) (“LUXTURNA® (Voretigene Neparvovec-Rzyl)”; Russell et al.). In a clinical trial, 31 subjects were enrolled, with ages ranging from 4 to 44 years. The patients either had a best visual acuity of 20/60, had a visual field of less than 20 degrees in any meridian, or both. 10 subjects were not treated with this gene replacement therapy; they served as the control group. The other 21 people received injections. In the control group, 30% of people (3 people) had a decrease in score by 1, while of the 21 with treatment, 0 people had a decrease in score from the MLMT test. In the treated group, 90% (19 people) had an increase in their MLMT score, while in the untreated group, only 40% (4 people) did. Also, without treatment, the highest increase in MLMT score was by 2, only by one person. With treatment, the highest increase in MLMT score

was by 4, also by one person. These results suggest that this therapy has the potential to help with vision restoration in people who have an IRD.

There are many possible side effects to LUXTURNA (“LUXTURNA® (Voretigene Neparvovec-Rzyl)”). In the clinical trial, 27 of 41 participants had an adverse reaction, ranging from mild to severe, in at least one eye. Of these 41 people, 46 of 81 eyes had some adverse reaction. The reactions ranged from increased eye pressure to deteriorating vision. LUXTURNA has potential side effects and cannot function for all IRDs, so more studies are still being conducted to create new treatments.

5.2 Partially Completed Clinical Studies Using Gene Replacement for IRDs

Since LUXTURNA cannot combat all IRDs, scientists are running clinical trials to test new alternatives. LUXTURNA only can treat IRDs that have mutations in the *RPE65* gene; however, many other genes can mutate to cause IRDs (“LUXTURNA® (Voretigene Neparvovec-Rzyl)”). Clinical trials are a phase that an investigational therapy has to undergo before requesting approval from the FDA (Commissioner, “Step 3”). They consist of four phases to ensure that the treatment has limited side effects, is safe, has a recommended dosage, and accomplishes its goal in humans. Phase I is to learn about side effects and the safety of treatments (*The Basics | National Institutes of Health (NIH)*). Phase I trials are tested on a group of 20-80 people. Next, during phase II trials, the effectiveness of the treatment is analyzed, and safety is further evaluated. In phase II trials, the number of participants grows to 100-300 people. Phase III trials are used to ensure the trial is effective, keep tracking side effects, and compare the treatment with other similar options. The group of people then becomes 1000-3000 people. Finally, during the last phase, phase IV, trials are approved by the FDA, and more information is collected using the general population. One phase II clinical trial that uses gene replacement is conducted by Nanoscope Therapeutics Inc.

The clinical trial by Nanoscope Therapeutics Inc. aimed to help with RP. MCO-010 is an investigational adeno-associated virus (a virus that is used in gene therapy to deliver genetic material), delivering multi-characteristic opsins (Nanoscope Therapeutics Inc., *Efficacy and Safety of MCO-010 Optogenetic Therapy in Adults With Retinitis Pigmentosa [RESTORE]*; Nanoscope Therapeutics Inc., *Non-Interventional Long Term Follow-up Study of Participants Previously Enrolled in the RESTORE Study*). The trial is called RESTORE. MCO-010 has the potential to work in most people with RP, no matter what gene causes it. It re-enables the eye to be sensitive to light. MCO-010 is a single-use, intravitreal injection. RESTORE is a phase 2b trial. There is an ongoing study with RESTORE that monitors their progress, a long-term study portion. RESTORE is not the only clinical trial being conducted; there are many different clinical trials being conducted to treat IRDs. See Table 1.

Reference number	Condition	Technology	Delivery method	Phase
NCT04945772 & NCT06162585	Retinitis pigmentosa	Insertion of MCO-010, an investigational adeno-associated virus	In vivo; intravitreal injections	2b
NCT05417126	Stargardt Disease	Insertion of vMCO-010	In vivo; intravitreal injections	2a
NCT05906953	Leber congenital amaurosis caused by mutations in RPE65	Insertion of HG004	In vivo; one-time insertion	1/2

Table 1: Clinical trials using gene replacement being conducted to treat IRDs in Phase 2

5.3 Pre-Clinical Studies to Treat IRDs Using Gene Replacement

Before being tested on humans, a treatment undergoes pre-clinical trials (Commissioner, “Step 2”). This tests to ensure that the treatment is not toxic.

One pre-clinical study using gene replacement was conducted in mice to treat Stargardt disease (Kong et al.). A lentiviral vector with a copy of the *ABCA4* gene was added into the retina on post-natal day 4 or 5. The mice were injected with 1µL of the vector into the eye. This was done to reduce the production of the toxic A2E pigment, which the mutated gene causes to be produced in excess. Additionally, to evaluate how long the treatment took to be in effect, the tissues were further examined at various time points. The results of the trial showed that 5-20% of the photoreceptors were able to obtain a copy of the correct *ABCA4* gene from the vector. Furthermore, the amount of A2E pigment was only 8-12pmol per eye in the treated mice, while it was near 27-39pmol in the untreated mice, highlighting the efficacy of the treatment. See Table 2 to reference more pre-clinical trials using gene replacement for IRDs.

DOI	Condition	Technology
https://doi.org/10.1007/s00439-016-1637-y	Retinitis pigmentosa	Adeno-associated virus gene vector to help with MERTK-related retinitis pigmentosa
https://doi.org/10.1089/hum.2019.297	Retinitis pigmentosa	Adeno-associated virus (rAAV) vector AGTC-501 to treat X-linked retinitis pigmentosa with mutations in the <i>RPGR</i> gene
https://doi.org/10.1038/gt.2008.78	Stargardt disease	Lentiviral vector injected into mice on post-natal day 4 or 5 into the retina with the <i>ABCA4</i> gene
https://doi.org/10.1167/iovs.13-11871	Stargardt disease	Equine infectious anemia virus-based lentiviral vector with a healthy copy of <i>ABCA4</i> gene

https://doi.org/10.1038/mt.2009.277	Leber congenital amaurosis	Injection of adeno-associated viral (AAV) vectors of <i>RPE65</i> (AAV2-hRPE65v2)

Table 2: Pre-clinical trials for IRDs using gene replacement

Gene replacement, even with its advantages over conventional treatments, does not fix the core of the problem, the mutated gene.

6.0 Gene Editing for IRDs: A One-Time Fix

While gene replacement does not eliminate the mutated gene, gene editing directly targets the mutation. With gene editing, a premature stop codon is caused, or the gene is directly fixed. The main gene editing technology is CRISPR. CRISPR technology is based on the processes in bacteria used to protect themselves from viruses (Shen et al.). Additionally, the administration of CRISPR does not require years of training, making it easier to implement. Specifically, CRISPR Cas9 nucleases are a promising option for gene editing treatments.

6.1 CRISPR Cas9 Nucleases for IRDs

CRISPR Cas9 nucleases are a new technology that directly fix a mutated gene or cause a premature stop codon. CRISPR Cas9 nucleases consist of two main parts: a guide RNA that matches the target gene and the Cas9 protein. The name CRISPR stands for clustered regularly interspaced palindromic repeats (Redman et al.). The Cas9 comes from the Cas9 protein, which causes a double-stranded break in DNA, allowing genes to be modified. The guide RNA directs the Cas9 protein to the correct site, where it cuts the DNA. There are two types of repair mechanisms following CRISPR Cas9 nuclease activity: non-homologous end-joining (NHEJ) and homology-directed repair (HDR).

With NHEJ, there is no homologous template; the two ends of the DNA attach after the double-stranded break, often causing deletions or insertions (indels). NHEJ is an error-prone mechanism of repair, in which a frameshift mutation is intended to ablate the target gene expression.

The other way for DNA to be repaired is through HDR. With HDR, a homologous template is inserted with the protein and guide RNA (Rein et al.). This homologous template goes in between the two cut ends, repairing the gene. HDR can only be performed when a cell is in its S or G2 phase; it is best for actively dividing cells. HDR takes longer than NHEJ, but it has a lower risk of errors because it uses a template. While HDR is overall a more controlled mechanism than NHEJ, it is not efficient in non-dividing cells. Therefore, NHEJ is oftentimes more suitable than HDR for therapeutic applications.

In 2023, the first treatment using CRISPR Cas9 nucleases was approved by the FDA, emphasizing its efficacy. It is called CASGEVY and relies on NHEJ (*Patient Information | CASGEVY® (Exagamglogene Autotemcel)*). CASGEVY is used to treat sickle cell disease and is done ex vivo.

CRISPR Cas9 nucleases enable the gene to be directly edited, differentiating them from gene replacement therapy, which solely adds a healthy copy. Many pre-clinical trials are being conducted evaluating CRISPR Cas9 nucleases' efficacy for different genes that cause IRDs.

An ongoing pre-clinical study for RP uses CRISPR Cas9 nucleases to combat the S3334ter mutation in the rhodopsin gene. (Bakondi et al.) For this treatment, a single subretinal injection of a plasmid containing a 20-nucleotide guide RNA, Cas9, and the mCherry red fluorescent protein is administered to rats on P0. This was immediately followed by electroporation to allow the treatment to enter the photoreceptor cells. On P4, the cells containing the mCherry protein were sorted, DNA was extracted, and the DNA was amplified by. In the amplified cells of the treated rats, 33-36% of the cells were found to have indels, and in the 8 predicted off-target edit spots, no indels were found. Furthermore, the treated rats had 8 layers of photoreceptors in the outer nuclear layer of the retina, while the untreated rats only had 1, emphasizing the efficacy of the treatment in preserving vision. Table 3 provides additional pre-clinical trials that use CRISPR Cas9 nucleases for IRDs.

DOI	Condition	Technology
10.1038/mt.2015.220	Retinitis pigmentosa	In vivo use of CRISPR-Cas9 nucleases targeting the S334ter mutation
10.1016/j.omtn.2023.05.008	Stargardt disease	CRISPR Cas9 nucleases to fix a mutation in the <i>ABCA4</i> gene in human induced pluripotent stem cells
10.1016/S1525-0016(16)34296-4	Leber congenital amaurosis	CRISPR Cas9 nucleases to fix the IVS26 c. 2991+1655 A>G mutation in the <i>CEP290</i> gene

Table 3: Pre-clinical trials using CRISPR Cas9 nucleases for IRDs

6.2 Base Editing as a Solution for IRDs

Many different pre-clinical studies that are attempting to treat IRDs use base editing. Base editing is another promising type of gene editing that uses CRISPR technology. Base editing focuses on changing single-nucleotides more accurately than CRISPR Cas9 nucleases (Porto et al.). Unlike CRISPR Cas9 nucleases, base editing does not rely on a double-stranded break; instead it chemically changes specific nucleotides. There are three types of base editing: CBEs, ABEs, and GBEs. CBEs help change G·C base pairs to A·T base pairs. CBEs consist of the nCas9 or dCas9 protein, gRNA, and cytidine deaminase (Bharat et al.). The gRNA directs the Cas9 protein to the DNA, and the cytidine deaminase converts cytosine to uracil, which is interpreted as thymine in DNA, leading to a nucleotide change. ABEs convert A·T base pairs to G·C base pairs. While CBEs use cytidine deaminase, ABEs use deoxyadenosine deaminase. GBEs are a newer type of base editing that cause transversions instead of only base transitions, allowing a change from C to G in mammalian cells (Zhao et al.). Base editing is used for its

accuracy in changing single-nucleotides without causing a double-stranded break. Many pre-clinical trials are being conducted using base editing for IRDs.

In a pre-clinical trial conducted to combat Stargardt disease, scientists tried to develop an adenine base editing strategy to fix the c.5882G>A mutation (Muller et al.). A 21-nucleotide guide RNA was selected, which goes to the target base of codon 1961 at the base editing window. They also edited an adjacent adenine, A8, which does not affect the amino acid sequence. They did this to use A8 as a marker. This allowed them to ensure that the base edit worked. Scientists found that an adenine base editor can correctly target the c.5882G>A mutation, as around 92% of the A8 gene was edited. It was also possible to edit rods, cones, and RPE cells, emphasizing its efficacy. 75% of the cones were edited, while 87% of the RPE cells were edited. Additionally, revealing its efficacy, editing was localized to the eye, and there were no significant off-target edits. Compared to the first-generation vector, the dose was reduced by 7.5x, resulting in less toxicity and less immune system stress. This made the trial safer overall. Even with all the results, there was one main limitation. Scientists were unable to measure the functionality of the treatment, as there was no phenotype of Stargardt disease even before the treatment. On the molecular level, the trial had great results; however, it is unknown what its exact effects are on vision. Table 4 describes more pre-clinical trials using base editing for IRDs.

DOI	Condition	Technology
https://doi.org/10.1038/s41591-024-03422-8	Stargardt disease	Base editing to correct the most common mutation of <i>ABCA4</i>
https://doi.org/10.1016/j.omtn.2022.11.021	Leber congenital amaurosis	Dual adeno-associated virus containing adenine base editors to treat mice with a retinal degeneration of 12, and particularly the <i>RPE65</i> gene
https://doi.org/10.1016/j.ymthe.2025.03.021	Retinitis pigmentosa	Adeno-associated virus vector with adenine base editors and a single guide RNA to target the T17M mutation
https://doi.org/10.3389/fgene.2021.814131	Stargardt disease	Analyzing data to find mutations and determining if they are editable with base editing

Table 4: Pre-clinical studies using CRISPR base editing to treat IRDs

6.3 Safety and Ethical Considerations of Gene Therapies

Even with their promise for treating IRDs and other genetic diseases, gene therapies have potential ethical and safety concerns. Harmful side effects, such as cataracts and retinal hemorrhaging, have been reported in a percentage of patients treated with LUXTRNA. Particularly, 57% of eyes experienced adverse effects; however, it was determined that the

benefit of helping people with certain mutations in the *RPE65* gene outweighed the risk of side effects (“LUXTURNA® (Voretigene Neparvovec-Rzyl)”).

Additionally, with gene editing technology, particularly CRISPR Cas9 nucleases, there is a chance that a non-target gene is edited, which could cause new problems to arise (Kalter et al.). A potential consequence is the development of cancer. This can occur due to an off-target edit that either prevents a gene from suppressing a tumor or begins promoting tumor formation. Even when there are no visible phenotypes, off-target edits can still cause harmful proteins (Wienert and Cromer).

Gene therapies are also very expensive, meaning they may not be accessible (Rueda et al.; “CRISPR & Ethics”; Raposo). LUXTURNA costs around \$425,000 per eye and \$850,000 for both eyes (“Spark’s Gene Therapy Price Tag”). As new gene editing technologies are developed, they may be more expensive than LUXTURNA and cost prices near or above the \$2.2 million that CASGEVY costs.

7.0 Discussion

This review focused on the treatment options for inherited retinal disorders and how they work. CRISPR base editing has proved the most promising for treating IRDs, as it has a lower risk of off-target edits than CRISPR Cas9 nucleases and is able to directly combat the mutation unlike gene replacement.

While progress is being made to help IRDs, there are still a very limited number of treatments. The majority of trials at the clinical stage use gene replacement rather than gene editing, which cannot fix the gene mutation because it only adds a healthy copy. This makes it unapplicable for many mutations, particularly if the mutated gene encodes a harmful protein. The harmful protein will still be produced, meaning that gene replacement was ineffective for this situation.

For the more than 250 genes that cause IRDs, only one treatment option is available; it only helps with the *RPE65* gene and is called LUXTURNA (“LUXTURNA® (Voretigene Neparvovec-Rzyl)”). This reveals that there are currently extremely limited options approved by the FDA. Furthermore, LUXTURNA is a gene replacement therapy, which has more limitations than gene editing. Different pre-clinical trials are being conducted with CRISPR Cas9 nucleases and CRISPR base editing to directly fix gene mutations. CRISPR base editing expands on CRISPR Cas9 nucleases technology, only needing to nick the DNA instead of causing a double-stranded break; this reduces the risk of off-target edits.

Off-target edits are a major safety concern of gene editing treatments, as the long-term effects are unknown. Additionally, other adverse effects are significant safety concerns; LUXTURNA has a 57% risk rate of adverse effects (“LUXTURNA® (Voretigene Neparvovec-Rzyl)”). A more ethical worry is the cost of these treatments, as LUXTURNA costs \$850,000 (“Spark’s Gene Therapy Price Tag”) and CASGEVY more than \$2.2 million (“CRISPR & Ethics”).

While potentially expensive, CRISPR Cas9 nucleases, CRISPR base editing, and gene replacement are currently three of the most developed gene therapies. A less developed gene therapy is prime editing. While it does not eliminate the price problem, prime editing is still a potential option for future gene therapies (Scholefield and Harrison). It uses a reverse transcriptase enzyme fused to the Cas9 nickase, and instead of a standard gRNA, it uses a pegRNA. pegRNA contains both an RNA template and a guide RNA. Additionally, prime editing causes a nick in the DNA instead of a double-stranded break, making it less likely to cause an off-target edit. Prime editing, to be able to be applied in the medical world, must be researched more; however, it still presents a promising new gene editing option.

Gene editing, particularly CRISPR, is expanding quickly compared to vaccines, emphasizing the promise of prime editing and other gene editing techniques being used in the near future. In 1993, Francisco Mojica characterized a CRISPR locus, and in 2005, he found that it is an adaptive immune system (“CRISPR Timeline”). Also in 2005, Alexander Bolotin discovered PAM sequences and the Cas9 protein. In 2013, Feng Zhang successfully used CRISPR Cas9 for genome editing in eukaryotic cells, particularly human and mouse cells. In 2023, CASGEVY was the first approved CRISPR Cas9 nucleases treatment (*Patient Information | CASGEVY® (Exagamglogene Autotemcel)*). It took 20 years from the discovery of CRISPR in bacteria to the approval of CASGEVY by the FDA. Compared to vaccines, CRISPR has been developing quickly. Since as early as 200 BCE, healthy people have been purposefully exposed to viruses to try to create immunity, the idea behind vaccines (*A Brief History of Vaccination*). In 1774, Benjamin Jesty injected cowpox into humans and found that it protected them from smallpox, confirming the ideas that had been used for almost 2000 years. Later, in 1796, James Phipps, an 8-year-old boy, was injected with cowpox by Edward Jenner. Phipps recovers and is later injected with smallpox, becoming the first person ever to be vaccinated. In 1872, Louis Pasteur created the first lab-produced vaccine for chickens. In comparison with CRISPR Cas9 nucleases, vaccines were being developed very slowly. Ideas similar to vaccines had been used since ancient times, and even after being proven by Edward Jenner, it took 76 years for the concept to be created in a laboratory. Much of this may have been due to technological limitations at the time; however, in comparison, CRISPR Cas9 nucleases are developing rapidly.

CRISPR Cas9 nucleases and other gene therapies will be able to help more than 5.5 Million people with IRDs (Hanany et al.). As gene therapies for IRDs advance, more will be known about these techniques as a whole, helping expand knowledge for other treatments. It is important to progress gene therapies for IRDs and other genetic disorders, and research should continue. These treatments offer ways to cure blindness that cannot be helped with any conventional treatments; gene therapies are the only fix.

8.0 Conclusion

Gene therapies are tools that, when used safely, can positively transform the medical world, helping people with IRDs and other genetic disorders. Currently, there are a limited number of treatments approved by the FDA, as most treatments are still in clinical or pre-clinical

phases. Even when treatments are available, they are still very expensive, limiting their accessibility. However, for IRDs as a whole, gene therapies have already begun to make a large impact on the lives of many people. 90% of patients in LUXTURNA trials greatly improved their test scores, emphasizing its efficacy (“LUXTURNA® (Voretigene Neparvovec-Rzyl)”). Furthermore, LUXTURNA can help with many different IRDs, as the *RPE65* gene mutations can cause a range of IRDs. Many clinical trials are being conducted, and as time passes, more treatments will be released, helping a significant number of people. These clinical and pre-clinical trials are being conducted to address different genetic disorders. While gene replacement is more promising than conventional treatments, CRISPR base editing and CRISPR Cas9 nucleases have greater long-term potential as they fix the gene mutation itself. Overall, the development of gene therapies has accelerated over the past few years, and significant progress has been made. Gene therapies have the potential to revolutionize the medical world as they can cure IRDs and other genetic disorders.

Works Cited

- A Brief History of Vaccination.
<https://www.who.int/news-room/spotlight/history-of-vaccination/a-brief-history-of-vaccination>. Accessed 11 Feb. 2026.
- Alberts B, et al. *Molecular Biology of the Cell*. 4th Edition, Garland Science, 2002.
- Bakondi, Benjamin, et al. “In Vivo CRISPR/Cas9 Gene Editing Corrects Retinal Dystrophy in the S334ter-3 Rat Model of Autosomal Dominant Retinitis Pigmentosa.” *Molecular Therapy*, vol. 24, no. 3, Mar. 2016, pp. 556–63. DOI.org (Crossref), <https://doi.org/10.1038/mt.2015.220>.
- Battu, Rajani, et al. “Newer Therapeutic Options for Inherited Retinal Diseases: Gene and Cell Replacement Therapy.” *Indian Journal of Ophthalmology*, vol. 70, no. 7, July 2022, pp. 2316–25. DOI.org (Crossref), https://doi.org/10.4103/ijo.IJO_82_22.
- Bharat, Sutar Suhas, et al. “Base Editing in Plants: Current Status and Challenges.” *The Crop Journal*, vol. 8, no. 3, 2020, pp. 384–95,
<https://doi.org/https://doi.org/10.1016/j.cj.2019.10.002>.
- Commissioner, Office of the. “Step 2: Preclinical Research.” FDA, Apr. 2019. www.fda.gov,
<https://www.fda.gov/patients/drug-development-process/step-2-preclinical-research>.
- Commissioner, Office of the. “Step 3: Clinical Research.” FDA, Nov. 2025. www.fda.gov,
<https://www.fda.gov/patients/drug-development-process/step-3-clinical-research>.
- “CRISPR & Ethics.” Innovative Genomics Institute (IGI),
<https://innovativegenomics.org/crisprpedia/crispr-ethics/>. Accessed 19 Nov. 2025.
- “CRISPR Timeline.” @broadinstitute, 25 Sept. 2015,
<https://www.broadinstitute.org/what-broad/areas-focus/project-spotlight/crispr-timeline>.
- Hanany, Mor, et al. “Worldwide Carrier Frequency and Genetic Prevalence of Autosomal Recessive Inherited Retinal Diseases.” *Proceedings of the National Academy of Sciences*, vol. 117, no. 5, Feb. 2020, pp. 2710–16. DOI.org (Crossref),
<https://doi.org/10.1073/pnas.1913179117>.
- Hwu, Wuh-Liang. “Gene Therapy for Ultrarare Diseases: A Geneticist’s Perspective.” *Journal of Biomedical Science*, vol. 31, no. 1, Dec. 2024, p. 79. link.springer.com,
<https://doi.org/10.1186/s12929-024-01070-1>.
- “Inherited Retinal Diseases.” ASGCT,
<https://patienteducation.asgct.org/understanding-cell-gene-therapy/conditions-treated/sensory-disorders/inherited-retinal-diseases>. Accessed 17 Nov. 2025.
- Inherited Retinal Disorders (IRD) | Boston Children’s Hospital.
<https://www.childrenshospital.org/conditions/inherited-retinal-disorders>. Accessed 17 Nov. 2025.
- Kalter, Nechama, et al. “Off-Target Effects in CRISPR-Cas Genome Editing for Human Therapeutics: Progress and Challenges.” *Molecular Therapy Nucleic Acids*, vol. 36, no. 3, Sept. 2025, p. 102636. ScienceDirect, <https://doi.org/10.1016/j.omtn.2025.102636>.
- Kong, J., et al. “Correction of the Disease Phenotype in the Mouse Model of Stargardt Disease

- by Lentiviral Gene Therapy.” *Gene Therapy*, vol. 15, no. 19, Oct. 2008, pp. 1311–20. DOI.org (Crossref), <https://doi.org/10.1038/gt.2008.78>.
- Leber Congenital Amaurosis: MedlinePlus Genetics. r. Accessed 17 Nov. 2025.
- “LUXTURNA® (Voretigene Neparvovec-Rzyl).” LUXTURNA® (Voretigene Neparvovec-Rzyl), <https://luxturna.com/>. Accessed 14 Nov. 2025.
- Muller, Alissa, et al. “High-Efficiency Base Editing in the Retina in Primates and Human Tissues.” *Nature Medicine*, vol. 31, no. 2, Feb. 2025, pp. 490–501. DOI.org (Crossref), <https://doi.org/10.1038/s41591-024-03422-8>.
- Nanoscope Therapeutics Inc. A Phase 2b Randomized, Double-Masked, Sham-Controlled, Study to Evaluate the Efficacy and Safety of Intravitreal Injection of MCO-010 Optogenetic Therapy in Adults With Retinitis Pigmentosa [RESTORE]. Clinical trial registration no. NCT04945772, clinicaltrials.gov, 20 Mar. 2024. [clinicaltrials.gov](https://clinicaltrials.gov/study/NCT04945772), <https://clinicaltrials.gov/study/NCT04945772>.
- Nanoscope Therapeutics Inc. Long Term Follow-up for Subjects Who Previously Participated in the NTXMCO-002 RESTORE Study. Clinical trial registration no. NCT06162585, clinicaltrials.gov, 21 Mar. 2025. [clinicaltrials.gov](https://clinicaltrials.gov/study/NCT06162585), <https://clinicaltrials.gov/study/NCT06162585>.
- Patient Information | CASGEVY® (Exagamglogene Autotemcel). <https://www.casgev.com/sickle-cell-disease>. Accessed 22 Nov. 2025.
- Porto, Elizabeth M., et al. “Base Editing: Advances and Therapeutic Opportunities.” *Nature Reviews Drug Discovery*, vol. 19, no. 12, Dec. 2020, pp. 839–59. DOI.org (Crossref), <https://doi.org/10.1038/s41573-020-0084-6>.
- Raposo, Vera Lucia. “The First Chinese Edited Babies: A Leap of Faith in Science.” *JBRA Assisted Reproduction*, vol. 23, no. 3, Aug. 2019, pp. 197–99. PubMed, <https://doi.org/10.5935/1518-0557.20190042>.
- Rare Genetic Diseases. <https://www.genome.gov/dna-day/15-ways/rare-genetic-diseases>. Accessed 14 Nov. 2025.
- Redman, Melody, et al. “What Is CRISPR/Cas9?” *Archives of Disease in Childhood. Education and Practice Edition*, vol. 101, no. 4, Aug. 2016, pp. 213–15. PubMed, <https://doi.org/10.1136/archdischild-2016-310459>.
- Rein, Lindsay A. M., et al. “Applications of Gene Editing Technologies to Cellular Therapies.” *Biology of Blood and Marrow Transplantation*, vol. 24, no. 8, Aug. 2018, pp. 1537–45, <https://doi.org/10.1016/j.bbmt.2018.03.021>.
- Retinitis Pigmentosa | National Eye Institute. <https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/retinitis-pigmentosa>. Accessed 27 Nov. 2025.
- Road, Foundation Fighting Blindness Phone:683-5555 6925 Oakland Mills, et al. “What Is Stargardt Disease?” Foundation Fighting Blindness, <https://www.fightingblindness.org/diseases/stargardt-disease>. Accessed 27 Nov. 2025.

- RPE65 Gene: MedlinePlus Genetics. <https://medlineplus.gov/genetics/gene/rpe65/>. Accessed 15 Nov. 2025.
- Rueda, Jon, et al. “Affordable Pricing of CRISPR Treatments Is a Pressing Ethical Imperative.” *The CRISPR Journal*, vol. 7, no. 5, Oct. 2024, pp. 220–26. DOI.org (Crossref), <https://doi.org/10.1089/crispr.2024.0042>.
- Russell, Stephen, et al. “Efficacy and Safety of Voretigene Neparvovec (AAV2-hRPE65v2) in Patients with RPE65-Mediated Inherited Retinal Dystrophy: A Randomised, Controlled, Open-Label, Phase 3 Trial.” *Lancet* [London, England], vol. 390, no. 10097, Aug. 2017, pp. 849–60. PubMed, [https://doi.org/10.1016/S0140-6736\(17\)31868-8](https://doi.org/10.1016/S0140-6736(17)31868-8).
- Scholefield, Janine, and Patrick T. Harrison. “Prime Editing – an Update on the Field.” *Gene Therapy*, vol. 28, nos. 7–8, Aug. 2021, pp. 396–401. DOI.org (Crossref), <https://doi.org/10.1038/s41434-021-00263-9>.
- Shen, Shengfu, et al. “CRISPR as a Strong Gene Editing Tool.” *BMB Reports*, vol. 50, no. 1, Jan. 2017, pp. 20–24. PubMed, <https://doi.org/10.5483/bmbrep.2017.50.1.128>.
- “Spark’s Gene Therapy Price Tag: \$850,000.” *Nature Biotechnology*, vol. 36, no. 2, Feb. 2018, pp. 122–122. DOI.org (Crossref), <https://doi.org/10.1038/nbt0218-122>.
- Tatour, Yasmin, and Tamar Ben-Yosef. “Syndromic Inherited Retinal Diseases: Genetic, Clinical and Diagnostic Aspects.” *Diagnostics*, vol. 10, no. 10, Oct. 2020, p. 779. DOI.org (Crossref), <https://doi.org/10.3390/diagnostics10100779>.
- The Basics | National Institutes of Health (NIH). 24 Apr. 2025, <https://www.nih.gov/health-information/nih-clinical-research-trials-you/basics>.
- Wienert, Beeke, and M. Kyle Cromer. “CRISPR Nuclease Off-Target Activity and Mitigation Strategies.” *Frontiers in Genome Editing*, vol. 4, Nov. 2022. Frontiers, <https://doi.org/10.3389/fgeed.2022.1050507>.
- Zhao, Dongdong, et al. “Glycosylase Base Editors Enable C-to-A and C-to-G Base Changes.” *Nature Biotechnology*, vol. 39, no. 1, Jan. 2021, pp. 35–40. DOI.org (Crossref), <https://doi.org/10.1038/s41587-020-0592-2>.

Bridging the Gaps: How AI Influencers Have Transformed Communication Strategies and Fulfilled Latent Desires of the Youth By Aaria Varma

Research question- What are the gaps in communication strategies that AI influencers have managed to fill, and how have these developments fulfilled certain latent desires of the youth?

Introduction

In recent years, the landscape of influence, marketing, and digital communication has undergone a remarkable transformation. Whereas once the figure of the influencer was always rooted in the flesh-and-blood being — human creators, celebrities, macro- or micro-influencers who share their personal lives, experiences and opinions — a new type of influencer has emerged: AI influencers (also called virtual, digital, or synthetic influencers). These are personas constructed via algorithms, 3D modelling, computer graphics, generative AI, and often managed by teams of designers, animators, marketers, and engineers rather than by one individual's unmediated life. ^[1] AI influencers (for instance, Lil Miquela, Shudu, Kyra, Naina, Aitana López and others) are not just novelties; they are increasingly pervasive actors in brand campaigns, digital culture, youth identity formation, and social media strategy. ^[2]

The phenomenon of AI influencers marks a convergence of several technological, cultural, and psychological trends. Advances in 3D rendering, CGI, motion capture, generative AI, voice cloning, and algorithmic content recommendation have made it possible to create virtual characters that look realistic (or intentionally stylised), appear to have personalities, interact with audiences, endorse products, and in many respects mimic human influencers. The youth, growing up in eras of digital nativity, AI, algorithmic feeds, curated self-presentation, and online communities, are particularly receptive to such personas. ^[3] For many young people, digital life has always been mediated; they are accustomed to avatars, filters, video game characters, even CGI celebrities, making the line between real and virtual less rigid. ^[4]

What exactly is meant by AI/Virtual Influencers

AI influencers are virtual personas – often computer-generated or hybrid (blended real and artificial) – that have social media profiles, followers, and engage in content creation typically associated with human influencers. ^[5] They may be entirely synthetic (CGI modeled, voice generated, animated) or they may be hybrid (for example employing a “body double,” using human models whose body or movement is used, then overlaid with virtual face or features). ^[6] They are controlled (scripted narratives, brand-aligned messaging), designed for consistency, and managed to maximize particular brand and communicative outcomes. ^[7]

How do they differ from ‘usual’ (human) influencers

Human influencers bring lived experience, emotional changeability, spontaneity, imperfections, authenticity (or perceived authenticity), personal identity, and personal narrative. AI influencers sacrifice or transform many of those traits. Their authenticity is constructed or

performative; their experiences may be entirely fictional or designed; their consistency, predictability, and controllability far exceed what most human influencers can maintain.^[8] Whereas human influencers often risk scandals, unpredictable behavior, inconsistencies, fatigue, and errors, virtual influencers are designed to minimize such risks and maximize brand alignment.^[9]

Examples and Characterizations

- **Lil Miquela:** One of the most famous virtual influencers, created by Brud. She has millions of followers, appears in fashion campaigns (Prada etc.), produces posts with aesthetic, personality, activism angles. She embodies a blend of fashion, identity politics, fantasy, and relatability. (Information about Lil Miquela is general in many articles; see “Virtual Influencers: New-age endorsement deals” listing her among AI influencers.^[10]
- **Shudu Gram:** Considered a “digital supermodel,” created by photographer Cameron-James Wilson in 2017. Her appearance is highly stylised, and she has been featured in high fashion campaigns. Her existence has sparked debate about representation, race, and realism.^[11]
- **Kyra (Kyraonig):** India’s first virtual influencer, created by Futr Studios. Launched in 2022. Kyra is visibly CGI (3D animated), involved in fashion/lifestyle/fitness content, has done multiple brand campaigns (e.g., Amazon Prime Video, boAt, Titan). She was designed using tools such as Unreal Engine, Maya etc.^[12]
- **Naina:** Created by Avtr Meta Labs, uses hybrid technology — body double, AI face-swap, AI motion capture, voice cloning etc. Combined with human interactions and real life situations to generate reels and content.^[13]
- **Aitana López:** A Spanish AI model created by The Clueless, known for her strong follower base on Instagram, income (in Euros), etc.^[14]

These examples differ in degree of realism, in strategy (full CGI vs hybrid), in how transparent they are about their artificiality, in what they promote, who controls them, and how they engage with their audiences.

History and Context

- Virtual/digital modeled characters have been around in some forms for years (CGI ads, avatars, video game influencers), but the recent wave of AI influencers has accelerated because of improvements in graphics tech, AI generative tools, social media’s appetite for new content, and brands’ search for controllability and novelty.
- In India, for example, Futr Studios co-founders noticed inspiration from virtual influencers abroad (Lil Miquela, Imma etc.) and saw a gap in the domestic market. They launched Kyra in 2022.^[15]

- As of 2023-2024, virtual influencer campaigns are reported to command fees that rival or even exceed those of human influencers with comparable reach. For example, Indian virtual influencers with 200,000-400,000 followers reportedly make ₹5-10 lakh per campaign, significantly higher than human influencers of similar following.^[16]
- Technological tools: 3D modeling tools like Maya, Blender, Unreal Engine; generative AI; voice cloning; motion capture; face swapping; AI motion capture technology. These tools allow creation of avatars or hybrids that can interact, move, speak, even appear in multiple locales virtually.^[17]

Problematizing / Statement of the Problem

Despite the increasing visibility and commercial success of AI influencers, there is much that remains underexamined: what exactly are the gaps in communication that human-led influencer strategies leave open, that AI influencers are able to fill? What latent psychological or emotional needs of youth are being met by these synthetic figures — needs that human influencers may fail to satisfy, or only partially satisfy? And how do youth experience authenticity, trust, fantasy, relatability, escapism, and self-identity in relation to AI influencers versus human influencers? These are important questions because the growing role of AI influencers could have implications for identity formation, media consumption habits, marketing ethics, and societal norms around what is “real” in digital life.

Major Argument / Thesis

In this paper, I will argue that AI influencers have become powerful communicative agents precisely because they bridge gaps left by human influencers—gaps in consistency, availability, controlled image, fantasy, and non-judgmental engagement—and in doing so, they fulfill latent desires among youth such as idealization, stability, curated relatability, escapism, and safe experimentation with identity. While AI influencers are not substitutes for authentic human experience, they offer complementary modes of interaction that human influencers seldom can sustain. To explore this thesis, the paper proceeds as follows: Section 1, titled “Imitation, Enchantment, and Longings,” examines how AI influencers can mimic human behavior and the psychological mechanisms, such as willing suspension of disbelief and parasocial relationships, that allow youth to be enamored by virtual personalities even when they are aware of their artificiality. This section includes case studies of AI influencers such as Lil Miquela, Shudu, and Kyra, highlighting what makes them particularly compelling and how they compare with human influencers in terms of perceived authenticity, emotional depth, and connection. Section 2, “Advantages, Disadvantages, and Implications for Marketing & Identity,” analyses how brands and marketers utilize AI influencers, exploring the studios behind them, the strategies employed, and the benefits they offer, including creative freedom, controlled narratives, cost efficiency, scalability, and risk management. It also addresses the potential drawbacks, including issues of authenticity, trust, transparency, ethical concerns, and the possible negative effects on youth identity, as well as the competitive implications for human influencers.

Section 1: Imitation, Enchantment, and Longings

Mimicry of Human Behaviour: How AI Influencers Approximate the Human

One of the first striking capabilities of AI influencers is their capacity to replicate many surface-features of human influencer behaviour: posting personal-style content, speaking in first person, showing “behind the scenes” stylized content, modeling aspirational lifestyles or fashion, speaking of feelings, causes, activism or identity issues. The appearance of personality is central: tone, voice, aesthetic choices, narratives of self (even fictional or curated), and interaction (likes, replies, comments).

For instance, Lil Miquela’s identity blends fashion, music, social commentary, activism, and emotional vulnerability; she posts about topics such as romance, social justice and mental health — all of which are classic tropes human influencers use to build intimacy and parasocial relationships. Though she is synthetic, she is constructed to feel like someone with a life story, struggles, opinions.

Shudu Gram also employs high fashion visuals, sculpture-like perfect form, aspirational beauty standards, runway aesthetic — traits human models embody — yet the perfection and stylization are amplified: image, lighting, presentation are controlled far more strictly than what most human influencers could sustain.

Kyra and Naina (India) similarly mirror human content tropes: fashion/lifestyle content, fitness, behind-the-scenes style, tutorials, collaborations, photo shoots. But they are unburdened by the unpredictability of human life — no fatigue, scandal, unscripted slipups — enabling consistent posting and consistent image.

The Youth’s Enchantment: Why even knowing it’s artificial doesn’t stop attachment

A paradox lies at the heart of the AI influencer phenomenon: many youths follow these avatars and virtual beings even when they are aware of their artificiality. Why does the knowledge of “fake” not fully block emotional investment?

- **Parasocial relationships:** Youth often form parasocial relationships (one-sided emotional attachments) with media figures. AI influencers extend this phenomenon into synthetic territory: the target of affection or admiration need not be human to become meaningful. The predictability and consistency of AI personalities help in sustaining these parasocial ties, sometimes more stably than a human whose mood, schedule or personal life may disrupt content flow.^[18]
- **Willing suspension of disbelief:** Youth are often willing to temporarily accept the fiction of AI influencers — to engage with them as if they were human, to imagine their feelings, even when they know they are virtual. This is similar to how people engage with fictional characters in novels, films, games. The aesthetic or experiential pleasure persists; authenticity shifts from being about reality to being about coherence, style, and narrative consistency.^[19]

- **Fantasy, idealization, escapism:** AI influencers can embody ideal forms: beauty without blemish, empowerment without visible conflict, activism without backlash, follower counts without burnout. For youth who face unpredictable real-life challenges, AI influencers offer safer spaces: for aspiration, for exploring identity (gender, style, race, aesthetics) in ways that may be unavailable or risky in real human contexts.^[20]
- **Relatability via curated image:** Paradoxically, because AI influencers are carefully designed, they can be both aspirational and relatable. They may express vulnerability, mental health concerns, or lifestyle suggestions in ways that are safe (because scripted), consistent, and controlled. Youth may see in them elements of what they want to become or what they wish to see — without the messiness of real life leaking in too much.^[21]

These mechanisms allow AI influencers to fill emotional, psychological, and communicative gaps that human influencers may leave.

Case Studies: What is so Likeable, Who Follows, What They Do Better

Lil Miquela

Lil Miquela (aka Miquela Sousa) is often considered the archetypal virtual influencer. She has millions of followers. Her content often merges fashion trends, music, activism, identity issues, and storytelling. She engages via Instagram, sometimes music, sometimes partnerships. Her identity as AI is either disclosed or understood, yet she occupies many of the spaces human influencers do: brand campaigns, style influencer, even social commentary. Because of her fictional narrative, she can address polarizing or idealized topics without real-life risk in reputation, legal liability or overspill.

Shudu Gram

Shudu is notable for her hyperreal visuals, stylised beauty, and fashion model aesthetics. Her image is about perfection in lighting, body, aesthetic, but it also raises issues (e.g. representation, race, authorship) because she is a Black-looking CGI model created by a white photographer; people question the ethics of likeness, representation, and the role of the creator. But for youth, Shudu offers a kind of fantasy: fashion model perfection, aspirational beauty, perhaps something unachievable but enjoyable to look at, aspire to, to emulate in small ways (style, makeup, imagery).

Kyra / Naina

In India, Kyra (Futr Studios) and Naina (Avtr Meta Labs) provide interesting contrast. Kyra is a fully virtual (3D animated), Naina hybrid. They create content in fashion, beauty, lifestyle, sometimes voice-enabled, some in multiple languages, sometimes crossing over real events using face swap or motion capture. Youth audiences in India appear drawn to them not just for novelty but for the possibility of something futuristic, aspirational, controllable, yet still

somewhat humanlike. The hybrid form gives additional cues of realism: real settings, human bodies with modified faces, visible content production, collaborations with real brands and celebrities.

What they do better than many human influencers

- **Consistency:** AI influencers can post regularly, maintain style, tone, and messaging without fatigue, scandal, off days.
- **Predictability and Safety:** Brands enjoy control over messaging, can avoid personal scandals, unpredictable behavior, or backlash that human influencers sometimes bring.
- **Creative Freedom & Fantasy:** They can appear in settings, do things, look perfect or stylised in ways that human influencers might avoid or cannot afford; they are not constrained by physical limitations (e.g. travel, tiredness) and ethical fragility in certain content.
- **Idealized but curated relatability:** They can choose to show vulnerability, personal-style content, mental health discourse etc., but always within bounds; they can satisfy yearning for connection without exposing messy or controversial real-life details.

Comparison to Human Influencers

While both human and AI influencers use similar platforms (Instagram, TikTok, YouTube etc.), tools (photography, video, editing, storytelling), and techniques (sponsored content, personal narrative, activism, idealization), the differences lie in:

Dimension	Human Influencers	AI Influencers
Authenticity / lived experience	Direct, lived, often messy, unpredictable	Constructed, curated, controllable; may feel “authentic” but not lived
Vulnerability	Real risk, real personal costs; authenticity sometimes comes at (or less risk) cost	Simulated, selective vulnerability; no real life downsides
Consistency & schedule	Subject to human constraints (health, mood, controversies etc.)	Can maintain high consistency, always “on”, always available
Creativity & spontaneity	Human spontaneity, cultural embeddedness, improvisation	Controlled spontaneity, but may be limited to what is scripted or programmed; risk of genericity or over-styled content
Risk & accountability	Real humans face consequences, public criticism, etc.	The creators/managers bear risk; but for the influencer/avatar itself, fault is more abstract; risks include backlash, disillusionment, ethical concerns

Algorithmic Nature and Sentience Question

AI influencers are fundamentally algorithmic. They are designed, scripted, their personalities programmed; their “voice” is mediated. There is no sentience (in the human or

biological sense), no genuine subjective experience (as far as we can tell). They produce content based on programming, design, feedback loops (engagement metrics etc.). The question of sentience is relevant not only philosophically but also in terms of how audiences perceive them: some may anthropomorphize, projecting feelings, personalities, but others may always retain an awareness that this is a construct.

This gap — between appearance of emotion/personality and the lack of lived experience — may be central to understanding how AI influencers fill certain needs: youth desire emotional connection, but perhaps do not always need the risk or messiness of real relationships or influencers.

Section 2: Advantages, Disadvantages, and Implications for Marketing & Identity

The creation and operation of AI influencers is increasingly handled by specialized studios and agencies that combine technological expertise with marketing acumen. In India, for example, Futr Studios launched Kyra in 2022,^[22] utilizing advanced 3D animation tools, Unreal Engine, and Maya to produce a fully digital persona capable of engaging local youth audiences. Similarly, Avtr Meta Labs developed Naina, a hybrid AI avatar that blends a human body double with face-swapping, motion capture, and voice cloning technologies. This multi-layered approach allows Naina to operate in multiple languages, participate in real-life events, and collaborate with brands on both virtual and offline campaigns.^[23] In Spain, The Clueless studio created Aitana López, a virtual model designed to target European fashion consumers, while global agencies like FFFACE.ME produce digital avatars for brands such as L'Oréal and Prada Beauty, integrating AI, AR, and digital fashion projects for marketing purposes. Across these examples, the creative teams decide not only the avatars' visual design and voice but also their narrative arcs, posting schedules, and content style, often calibrating these choices based on audience engagement analytics.

The advantages of using AI influencers for brands are both creative and practical. From a narrative perspective, AI influencers provide unprecedented creative freedom. Unlike human influencers, whose behaviors and schedules are subject to real-world constraints, digital avatars can maintain a controlled and consistent public persona. Brands and studios can script personalities that perfectly align with marketing objectives while pushing boundaries in aesthetic presentation or fantasy-based content that would be difficult for human influencers to sustain. For instance, AI avatars can be portrayed performing extraordinary feats or inhabiting highly stylized environments, allowing youth audiences to engage in aspirational identity play and immersive fantasy experiences.^[24]

Beyond narrative control, AI influencers offer economic and operational advantages. While the initial creation of high-quality avatars—including 3D modeling, motion capture, and rendering—can be expensive, the long-term cost efficiency is considerable. Once constructed, these virtual influencers can generate vast amounts of content with minimal incremental cost, avoiding human constraints such as travel, rest, or illness.^[25] In India, some virtual influencers command campaign fees comparable to or exceeding those of human influencers with similar

followings, indicating that brands perceive their value not only in content production but also in their reliability and controlled presentation.^[26]

Reliability and consistency are key factors in the appeal of AI influencers to both brands and young audiences. Unlike human influencers, AI avatars do not experience off days or become embroiled in scandals, allowing brands to maintain a coherent public image across global campaigns. Hybrid AI avatars such as Naina, capable of operating in multiple languages, are particularly useful for multinational brands seeking scalable engagement. For young audiences, this consistency creates a predictable and safe space for interaction, free from the interpersonal drama that often accompanies human influencers. Moreover, these avatars satisfy latent desires for idealization and aesthetic perfection, enabling youths to explore identity experimentation and aspirational narratives in a controlled, low-risk environment.

However, the use of AI influencers is not without its challenges and ethical implications. A primary concern is authenticity. Although avatars can simulate vulnerability or emotional expression, they lack lived experience, and some followers may perceive them as inherently less credible than human influencers. Studies indicate that while engagement metrics may be high, trust in AI influencers often lags behind trust in real-life counterparts, particularly when audiences value expertise, personal history, or genuine emotional disclosure.⁵ This issue is compounded by the “uncanny valley” effect, in which digital avatars that are nearly—but not perfectly—human-like can provoke discomfort or alienation. Awareness of their artificiality may trigger algorithm aversion among certain users, who resist algorithmic decision-making and perceive AI content as manipulative.^[27]

Ethical concerns also extend to intellectual property, representation, and transparency. The case of Shudu Gram, a Black-appearing digital supermodel created by a white artist, sparked debates around cultural appropriation and the commodification of racial identities. Similar questions arise regarding disclosure: audiences must be made aware that the influencer is virtual to avoid misleading them. Moreover, AI influencers can crowd out smaller human creators, as brands increasingly favor avatars that are controllable, always on-brand, and capable of generating highly polished content. This trend may inadvertently raise content production standards across the industry, increasing pressure on human influencers and potentially affecting mental health and authenticity in the creator economy.^[28]

Finally, the technical and financial barriers to creating high-quality AI influencers can be significant. Expertise in graphic design, motion capture, voice synthesis, and content strategy is required, potentially excluding smaller studios or independent creators from participation. Additionally, public backlash is a possibility: if audiences perceive deception or excessive manipulation, they may develop distrust toward both the influencer and the associated brand. While research demonstrates that attractiveness and entertainment value drive engagement with AI influencers, such content does not necessarily translate to purchase intention unless the avatar is perceived as credible and informative.

In summary, AI influencers offer brands creative freedom, operational efficiency, reliability, and the ability to engage young audiences’ desires for fantasy, identity exploration,

and idealized interaction. At the same time, they raise significant ethical, cultural, and relational challenges, including concerns about authenticity, representation, emotional depth, and market effects on human influencers. Understanding these trade-offs is crucial for marketers, cultural commentators, and researchers exploring the implications of AI-driven media on youth identity and engagement.^[29]

Implications for Youth Identity, Self-Perception, and Digital Interaction

- **Identity experimentation and aspiration:** AI influencers have introduced new avenues for youth to explore and express their identities. These digital personas offer a canvas for experimenting with various styles, beliefs, and affiliations without the immediate risks associated with real-life social interactions. This virtual experimentation allows for a form of identity play that can be both liberating and instructive, providing a safe space for youth to navigate aspects of gender, aesthetics, and even activism.^[30]
- **Media literacy & perception of ‘real’ vs ‘performance’:** However, this freedom comes with complexities. The line between reality and performance becomes increasingly blurred as AI influencers craft narratives that mimic human emotions and experiences. This simulation can lead to confusion among young audiences about what constitutes authenticity. While some may develop a keen awareness of these constructed identities, others might struggle to differentiate between genuine and curated personas, leading to skepticism or a sense of emotional detachment.^[31]
- **Emotional safety vs emotional numbness:** Moreover, the emotional dynamics of interacting with AI influencers present both opportunities and challenges. For some, these digital figures offer consistent and non-judgmental companionship, serving as a source of comfort and support. Yet, the knowledge that these influencers are algorithmically generated can create a sense of emotional distance, limiting the depth of connection and potentially leading to feelings of isolation.^[32]
- **Influence on consumption behaviour and desires:** The impact of AI influencers extends to youth consumption behaviors and aspirations. By promoting idealized aesthetics and lifestyles, these digital personas can shape desires and expectations, sometimes reinforcing unattainable standards of beauty and success. This can contribute to issues like self-objectification and body image concerns. Conversely, when AI influencers advocate for inclusivity, diversity, and sustainability, they have the potential to positively influence youth ideals and encourage more equitable perspectives.^[33]
- **Shifting marketing norms and ethics:** As AI influencers become more prevalent, they also influence marketing norms and ethical considerations. The integration of these digital figures into advertising campaigns raises questions about transparency, representation, and the ownership of digital identities. Youth, as both consumers and creators, are at the forefront of engaging with these evolving norms, navigating what they accept and what they challenge in the digital landscape.^[34]

The rise of AI influencers represents a watershed moment in the history of communication, digital marketing, and youth culture. What began as a speculative experiment in blending artificial intelligence with social media aesthetics has now grown into a global industry that redefines the meaning of influence. Throughout this paper, I have argued that AI influencers not only fill the gaps left behind by human influencers—such as inconsistency, susceptibility to scandals, and limited creative reach—but also fulfill certain latent desires among young people for perfection, fantasy, non-judgmental companionship, and uninterrupted engagement. By examining case studies such as Lil Miquela, Shudu, and Kyra, it became evident that these synthetic personas have succeeded in capturing the imagination of youth audiences by offering both relatability and aspiration in curated, algorithmically tailored ways.

At the same time, this paper demonstrated that AI influencers are not simply tools of entertainment or fantasy; they are carefully orchestrated marketing assets controlled by studios and brands. Their existence highlights the increasing corporatization of identity and communication, where narratives are engineered to resonate with youth values such as diversity, activism, and futurism—yet always within the controlled boundaries of a marketable persona. The advantages of AI influencers, from cost-efficiency and availability to the elimination of human error, undeniably make them a valuable tool for businesses navigating an oversaturated influencer economy. However, these benefits must be weighed against the disadvantages: the erosion of authenticity, the absence of lived human experience, ethical concerns about transparency, and the potential marginalization of human influencers who cannot compete with algorithmic perfection.

Despite the insights offered, this study is not without its limitations. First, AI influencers are still a relatively new phenomenon, and the available literature is sparse and evolving, which makes long-term predictions tentative at best. Second, this paper primarily examined well-known global AI influencers, which may not fully capture the diversity of digital markets across regions. Finally, while surveys and interviews reveal valuable insights into youth perceptions, they cannot encompass the entire range of complex emotional and psychological responses that young audiences may have toward synthetic personalities.

Nevertheless, the strength of the argument remains intact: AI influencers are more than a passing trend; they are an emerging force reshaping the very fabric of digital communication. Their ability to blend fantasy with marketing, while simultaneously offering youth audiences a sense of controlled engagement, marks a profound shift in how influence is conceptualized and consumed. Looking ahead, future research must focus on the ethical frameworks governing AI-driven communication, particularly around issues of transparency, accountability, and the psychological impact on impressionable audiences. Just as human influencers once redefined celebrity culture, AI influencers are redefining the future of mediated relationships—challenging us to rethink what authenticity, trust, and influence mean in an age where reality itself can be simulated.

Works Cited

- [1] BuiltIn, "What Are AI Influencers?" BuiltIn. Accessed September 18, 2025.
<https://builtin.com/articles/ai-influencer>.
- [2] Influencity, "AI Influencers: How Virtual Personalities are Shaping the Future," Influencity Blog, January 3, 2025, <https://influencity.com/blog/en/ai-influencer>.
- [3] Jayasingh, S., A. Sivakumar, and A. A. Vanathaiyan, "Artificial Intelligence Influencers' Credibility Effect on Consumer Engagement and Purchase Intention," *Journal of Theoretical and Applied Electronic Commerce Research* 20, no. 1 (2025): 17, <https://doi.org/10.3390/jtaer20010017>.
- [4] Shen, Zhen, "Shall Brands Create Their Own Virtual Influencers? A Study of Customer-Brand Engagement with Virtual Influencers," *Humanities and Social Sciences Communications* (2024), <https://doi.org/10.1057/s41599-024-02698-y>.
- [5] Wikipedia, "Shudu Gram," last modified August 2024, https://en.wikipedia.org/wiki/Shudu_Gram.
- [6] Wikipedia, "Kyra (Virtual Influencer)," last modified July 2024, https://en.wikipedia.org/wiki/Kyra_%28virtual_influencer%29.
- [7] Abidin, Crystal. *Internet Celebrity: Understanding Fame Online*. Bingley: Emerald Publishing, 2018.
- [8] Moustakas, John. "Lil Miquela: The Digital Avatar Changing the Face of Influencer Marketing." *Forbes*, May 2023.
<https://www.forbes.com/sites/johnmoustakas/2023/05/17/lil-miquela-the-digital-avatar>.
- [9] Vincent, James. "How Virtual Influencer Lil Miquela Is Changing Activism." *The Verge*, June 2021. <https://www.theverge.com/2021/6/14/lil-miquela-activism>.
- [10] BuiltIn. "What Are AI Influencers?" BuiltIn. Accessed September 18, 2025.
<https://builtin.com/articles/ai-influencer>.
- [11] Roose, Kevin. "The Rise of Shudu, the World's First Digital Supermodel." *New York Times*, April 2020.
- [13] Fessler, Leah. "Shudu, the World's First Digital Supermodel, Sparks Debate about Race and Representation." *Quartz*, March 2018.
<https://qz.com/1224180/shudu-the-worlds-first-digital-supermodel-sparks-debate>.
- [14] Srivastava, Radhika. "Kyra: India's First Virtual Influencer Is Already Winning Over Brands." *Economic Times*, July 2022.
- [15] Kadakia, Pankti Mehta. "What Is Driving the AI Influencer Trend?" *Forbesindia.com*, *ForbesIndia*, 23 Oct. 2024,
www.forbesindia.com/article/digital-stars-2024/what-is-driving-the-ai-influencer-trend/94464/1?utm_source=chatgpt.com. Accessed 21 Sept. 2025.
- [16] Himanshi Lohchab. "New-Age Endorsement Deals: Virtual Influencers Now Making Real Money." *The Economic Times*, *Economic Times*, 2023,
economictimes.indiatimes.com/tech/technology/new-age-endorsement-deals-virtual-influencers-now-making-real-money/articleshow/105571845.cms?from=mdr&utm_source.
- [18] Srivastava, Radhika. "Kyra: India's First Virtual Influencer Is Already Winning Over Brands." *Economic Times*, July 2022.

- [19] Influencer Marketing Hub. “AI, Digital Clones, and Virtual Influencers: The New Face of Fashion Marketing.” Influencer Marketing Hub, June 25, 2025.
<https://influencermarketinghub.com/ai-digital-clones-and-virtual-influencers-the-new-face-of-fashion-marketing/>.
- [20] Yu, J., et al. “Artificial Intelligence-Generated Virtual Influencer (AIGVI).” *Technology in Society* (2024). <https://www.sciencedirect.com/science/article/pii/S0969698923003119>
- [21] BuiltIn, “What Are AI Influencers?” BuiltIn. Accessed September 18, 2025.
<https://builtin.com/articles/ai-influencer>
- [22] Srivastava, Radhika. “Kyra: India’s First Virtual Influencer Is Already Winning Over Brands.” *Economic Times*, July 2022.
- [23] Influencer Marketing Hub. “AI, Digital Clones, and Virtual Influencers: The New Face of Fashion Marketing.” Influencer Marketing Hub, June 25, 2025.
<https://influencermarketinghub.com/ai-digital-clones-and-virtual-influencers-the-new-face-of-fashion-marketing/>
- [24] Yu, J., et al. “Artificial Intelligence-Generated Virtual Influencer (AIGVI).” *Technology in Society* (2024). <https://www.sciencedirect.com/science/article/pii/S0969698923003119>
- [25] BuiltIn, “What Are AI Influencers?” BuiltIn. Accessed September 18, 2025.
<https://builtin.com/articles/ai-influencer>
- [26] Jayasingh, S., A. Sivakumar, and A. A. Vanathaiyan, “Artificial Intelligence Influencers’ Credibility Effect on Consumer Engagement and Purchase Intention,” *Journal of Theoretical and Applied Electronic Commerce Research* 20, no. 1 (2025): 17,
<https://doi.org/10.3390/jtaer20010017>
- [27] Harchandani, Chintan, Anurag Singh Chandrawat, and Akshita Upadhyay. “The Impact of Artificial Intelligence on Trustworthiness and Authenticity in Green Influencer Marketing.” *Integrated Journal for Research in Arts and Humanities* 5, no. 2 (2025): 41–46. <https://doi.org/10.55544/ijrah.5.2.7>
- [28] Fessler, Leah. “Shudu, the World’s First Digital Supermodel, Sparks Debate about Race and Representation.” *Quartz*, March 2018.
<https://qz.com/1224180/shudu-the-worlds-first-digital-supermodel-sparks-debate>.
- [29] Shen, Zhen, “Shall Brands Create Their Own Virtual Influencers? A Study of Customer-Brand Engagement with Virtual Influencers,” *Humanities and Social Sciences Communications* (2024), <https://doi.org/10.1057/s41599-024-02698-y>
- [30] The Social Institute, “A.I. Identity: How A.I. Is Impacting Students’ Self-Image, Relationships, and Online Searches,” accessed October 3, 2025,
<https://thesocialinstitute.com/blog/a-i-identity-how-a-i-is-impacting-students-self-image-relationships-and-online-searches/>
- [31] Jose, M., “Reality, Performance, and Virtual Influence: How AI Influencers Shape Youth Perception,” *Journal of Information Science* 51, no. 4 (2025): 653–668,
<https://journals.sagepub.com/doi/10.1177/01655515251330605>

- [32] Del Valle, R., “AI-Driven Identity Confusion: Adolescents’ Perspective,” LinkedIn, March 2025,
<https://www.linkedin.com/pulse/ai-driven-identity-confusion-adolescents-perspective-del-valle-bn4re/>
- [33] US Therapy, “AI Influencers: Opportunities and Risks for Youth,” accessed October 3, 2025,
<https://us-therapy.sg/insights/ai-influencers/>.
- [34] Jose, P., “Ethics, Marketing, and Virtual Personas: Youth Engagement with AI Influencers,”
Journal of Marketing Communications 31, no. 6 (2025): 1123–1140,
<https://journals.sagepub.com/doi/10.1177/026666669251374658>

Finding Resilience: Understanding the Psychosocial Determinants of Stress and Coping Strategies among Indian Boarding School Female Students By Evania Sahu

Abstract

This investigation reflects the findings of exploring stress and coping among students studying in an Indian residential school. The study assessed perceived stress, resilience, coping strategies, and emotion regulation among 25 female students aged 15 to 17 years. Data was acquired using standardised questionnaires such as the Perceived Stress Scale (PSS), Connor-Davidson Resilience Scale (CD-RISC-25), Emotion Regulation Questionnaire (ERQ), and Brief Coping Orientation to Problems Experienced Inventory (Brief-COPE), and descriptive statistical analysis was used to evaluate the results. The findings showed moderate levels of stress, resilience, and coping. Participants used cognitive reappraisal more than expressive suppression, indicating adaptive emotional management and coping.

Keywords Perceived stress, Resilience, Emotion regulation, Cognitive reappraisal, Expressive suppression, Adaptive coping strategies, Problem-focused coping

1. Introduction

Away from home and immersed in highly structured environments, within the walls of Indian boarding schools, adolescence is shaped by distance from family, academic expectations, and the constant demand to adapt. In these places, young adults not only learn academic lessons but also how to manage stress long before adulthood.

For every individual, adolescence is a crucial period of their overall development, marked by “emotional, mental, physical”, and social changes (World Health Organization). This period of time makes young people extremely prone to stress and anxiety due to academic pressure, relationships with their peers; building their identity, and adapting to different expectations of society and also those close to them (Jiang et al.). These pressures are often rendered harsher for boarding school students by persistent splits in families, strict rules and regulations, and restricted freedom for oneself. Boarding school students may face particular psychosocial challenges that have an extensive effect on their mental wellness, where academic performance is usually regarded as the main indicator of accomplishment.

Stress among young adults is a widely investigated and researched subject, with most studies claiming that chronic stress can create a significant negative impact on “emotional regulation, academic performance, athletic capabilities, self-esteem, and other sociopsychological” factors (Nations) (Kumar et al.). According to the "Transactional model of stress and coping", established by Richard S. Lazarus and Susan Folkman in 1984, this serves as a psychological framework for investigating how people interpret and react to stresses in the world around them (Kivak). According to this concept, stress is caused by interactions among people and their immediate environment rather than outside factors. This framework is significantly relatable to boarding school environments, where students must constantly adjust

to firm routines, demanding academic environments, and social hierarchies, making them interact with different individuals, environments, and challenges (Zhong et al.). Although global research has explored adolescent stress frequently, there has been minimal investigation into Indian residential school groups, despite their becoming more prevalent.

Coping methods and strategies are a crucial element in determining how adolescents manage stress and grow as more resilient individuals. Coping methods are typically divided into two types: problem-focused strategies, which try to tackle the root cause of stress, and emotion-focused strategies, which manage emotional reactions. According to research, effective coping methods—such as accessing social assistance, behavioural redirection, and problem-solving—are linked to improved mental wellness, whereas inefficient tactics, such as denial and suppressing feelings, can worsen stress (Rodrigues and Morouço). Boarding school students often face constraints in seeking their preferred coping mechanisms, particularly a lack of family support, rendering investigating their coping patterns significantly important.

Personality traits, emotional management abilities, perceived power, and confidence are all significant indicators of how young adults feel and respond to stress. Studies claim that traits such as resilience, optimism, self-confidence, and emotional maturity serve as protective tools, whereas poor emotional regulation and high neuroticism increase vulnerability to stress (Uliaszek et al.). In a more society-oriented culture like India, the psychosocial aspects are, moreover, shaped by societal and cultural norms, emphasising adherence to discipline, respect for authority and dominance of parents over their child's future. This mindset discourages emotional expression and help-seeking behaviours among youth. These aspects of culture can have an impact on how stress is perceived and ways to cope in boarding schools of India.

Social or peer assistance bodies act as some of the most consistent and powerful guards against stress. Support from family, peers, and teachers has been observed to improve emotional well-being, decrease anxiety, and enhance resilience in young adults. For those individuals who reside in boarding schools, peer relationships and teacher-student interactions frequently substitute for parental guidance, making their level of support and availability particularly vital. Positive peer relationships can provide a sense of belonging and understanding, whilst supportive teachers can act as mentors and emotional anchors (Butler et al.). In contrast, circumstances characterised by authoritative regulation, competitiveness among peers, or a lack of emotional sensitivity could increase stress and interfere with healthy coping.

Despite current research on adolescents and stress, there exists a significant gap in research that particularly investigates the relationship between stress, coping methods, and psychosocial factors amongst Indian boarding educational institution students. Most studies focus on suburban day students or take a solely academic approach to adolescent stress, ignoring students' daily realities and emotional expressions in residential school surroundings. Understanding these factors is critical when developing mental health treatment programmes, policies regarding education, and support systems that are suited to the requirements of these individuals.

This study seeks to bridge this gap by investigating the psychological causes of stress

and coping methods among Indian residential school students.

2. Methodology

2.1 Objectives

- a. To explore the experiences of stress and coping strategies among Indian boarding school youth.
- b. To identify the psychosocial determinants that influence stress and coping strategies among Indian boarding school youth.

2.2 Research Questions

- a. What are the common sources of stress experienced by Indian boarding school youth?
- b. How do Indian boarding school youth cope with stress, and what coping strategies are most effective?
- c. What psychosocial determinants (particularly emotional regulation) influence stress and coping strategies among Indian boarding school youth?

2.3 Hypothesis

Based on existing literature, it is hypothesized that Indian boarding school youth who have effective emotional regulation and adaptive coping strategies will experience lower levels of stress and higher levels of resilience.

For this particular study, a quantitative research design was utilized to assess the perceived stress, resilience, coping strategies, and emotional regulation among Indian boarding school students. The data collection took place by composing four standardised questionnaires, which were:

- Emotion Regulation Questionnaire (Gross, J.J., & John, O.P. (2003)).
- Perceived Stress Scale (Sheldon Cohen & Gail Williamson, 1988)
- CD-RISC-25- The Connor-Davidson Resilience Scale (*Kathryn M. Connor & Jonathan R.T. Davidson (2003)*)
- (Brief-COPE)- Brief - Coping Orientation to Problems Experienced Inventory (*Charles S. Carver (1997)*)

These questionnaires were compiled into a single form. The scoring guidelines and the scales were not disclosed or shared with the participants in order to reduce bias. Participation was entirely optional, and anonymity remained intact throughout the study to protect privacy and to foster truthful responses. The sample consisted of 25 female students between the ages of 15 and 17, all of whom were currently studying in a boarding school in India. The study was conducted at Stonehill International Boarding School in India. The participants were all part of the same school, so they had to share similar routines as a control variable. This study plan provides a comprehensive framework for exploring the psychosocial determinants of stress and

coping strategies among Indian boarding school youth. The findings can inform interventions and support systems to promote resilience and well-being among this population.

2.4 Ethical Considerations

- a. Informed consent was obtained from participants and their parents/guardians.
- b. Confidentiality and anonymity was ensured throughout the study.
- c. Participants were free to withdraw from the study at any time.

2.5 Data Analysis

Data (N=25) was analyzed using instructions as exactly given in the manuals of the respective standardized questionnaires. Descriptive statistics (mean, standard deviation and highest and lowest scores) were computed to examine stress levels, resilience, emotion regulation, and coping strategies. These results are visually represented on tables and bar graphs (Table 1 and Figures 1 to 4) to demonstrate patterns and relationships between variables.

3. Results

Descriptive statistics were calculated to examine the levels of “stress, resilience, emotion regulation, and coping strategies” among the sample of 25 Indian boarding school students.

Table 1: Descriptive Statistics of Study Variables (N = 25)

Variable	Mean	SD	Min	Max
Brief-COPE – Coping strategies	4.86	1.82	2	10
PSS – Perceived Stress Score	23.52	5.03	17	33
CD-RISC-25 – Resilience (<i>Connor-Davidson Resilience Scale (CD-RISC) © Manual</i>)	38.52	9.18	20	55
ERQ – Emotion Regulation Questionnaire	27	4.4	19	23

3.1 Perceived Stress

Perceived stress was measured using the Perceived Stress Scale (Sheldon Cohen, Tom Kamarck, and Robin Mermelstein, 1983). Scores ranged from 17 to 33, indicating variability in stress levels among participants. Mean stress score: 23.5. Based on conventional interpretation guidelines for the scale, the majority of participants fell within the moderate stress category, suggesting that boarding school students experience noticeable but manageable levels of psychological stress.

As seen in figure 1, the perceived stress depletes in the mid-range when compared to the other variables, indicating neither very high nor low levels of stress among the students. Even in Figure 3, it is illustrated that perceived stress scores are lower compared to the resilience scores, suggesting an inverse relationship between the two variables.

3.2 Resilience

Resilience was assessed using the Connor–Davidson Resilience Scale (*Davidson, 2018*).

Scores ranged from 20 to 55, with a mean score of 38.5. This indicates moderate resilience levels within the sample, suggesting that while students face stressors associated with boarding school life, they also demonstrate psychological capacities to adapt to challenging circumstances.

As seen in Figure 1, resilience has the highest average score among all measured variables. Even more, figure 3 clearly shows that resilience scores are substantially higher than perceived stress, highlighting the strong adaptive capacity of the participants despite the obstacles.

3.3 Emotion Regulation

Emotional regulation was measured using the *ERQ – Emotion Regulation questionnaire* – formally developed by James J. Gross and Oliver P. John in 2003. This scale is used to assess cognitive reappraisal and expressive suppression.

a. Cognitive Reappraisal

Cognitive reappraisal suggests the ability to reinterpret emotions in order to change their emotional and mental impact. In this investigation, the ERQ scores range from 19 to 33, with a mean score of 26.96 and a standard deviation of 4.46. These findings imply that those surveyed used cognitive reappraisal to moderately control their emotions. It is evident from figure 2 that cognitive reappraisal has the highest mean score among emotional regulation techniques, indicating that participants often rely on adaptive reinterpretation strategies to manage emotions.

b. Expressive suppression

'Expressive suppression' refers to the ability to control the expression of emotions externally. The scores for expressive suppression ranged from 8 to 25, with the average of 15.68 and a standard deviation of 4.60. This suggests that those surveyed had a moderate inclination to restrict or repress their emotional expressions in order to regulate their emotions.

3.4 Coping strategies

Coping strategies were assessed using the Brief-COPE (Brief Coping Orientation to Problems Experienced) inventory, which was innovated by Charles S. Carver and was used to measure and evaluate the different ways individuals respond to stress and obstacles in life. In the study, coping strategy scores ranged from 2 to 10. Mean stress score: 4.86. Standard deviation: 1.82

a. Adaptive coping strategies

'Adaptive coping strategies' refers to the different constructive strategies and methods utilised to manage stress, such as acceptance, seeking social or professional support, planning,

and positivity. In this investigation, the participants showed different kinds of adaptive coping, suggesting that students reportedly employed various techniques to cope with the challenges of residential school life.

As shown in Figure 4, there is variation among coping methods. Acceptance, preparing, and seeking help have greater mean scores than maladaptive methods like substance use and self-harm. This shows that individuals are more likely to use adaptive coping techniques.

b. Problem-focused coping

Problem-focused coping implies actively dealing with the root cause of stress by employing tactics such as planning and immediate action to overcome obstacles. The findings show that students engaged in problem-focused coping methods, indicating that many made an effort to handle stressful events proactively and in a solution-focused way.

In figure 4, strategies like planning and adaptive coping demonstrate moderates to high scores, supporting the presence of problem-focused coping among participants.

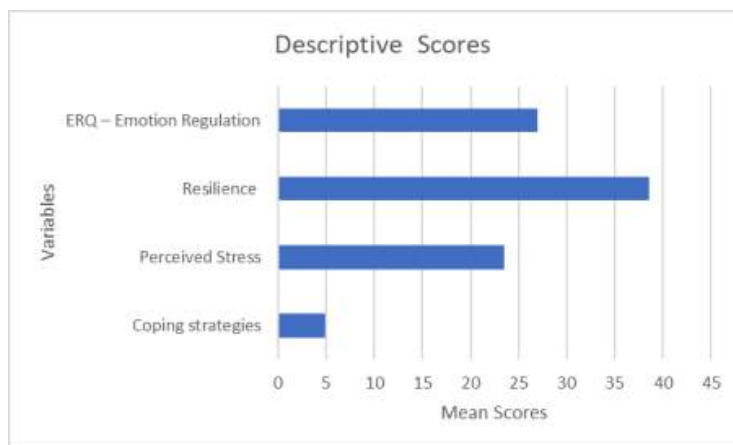


Fig 1: Relationship between Stress, Resilience, Coping strategies, and Emotion Regulation among Indian Boarding School Youth (N = 25)

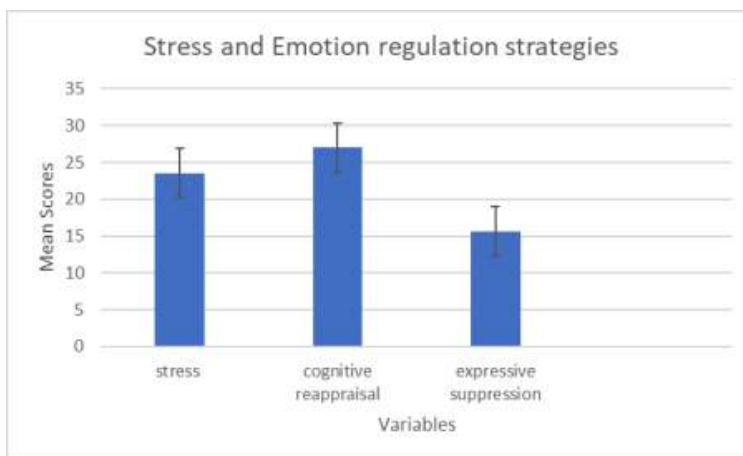


Fig 2: Relationship between Stress and Emotion Regulation strategies among Indian Boarding School Youth (N =

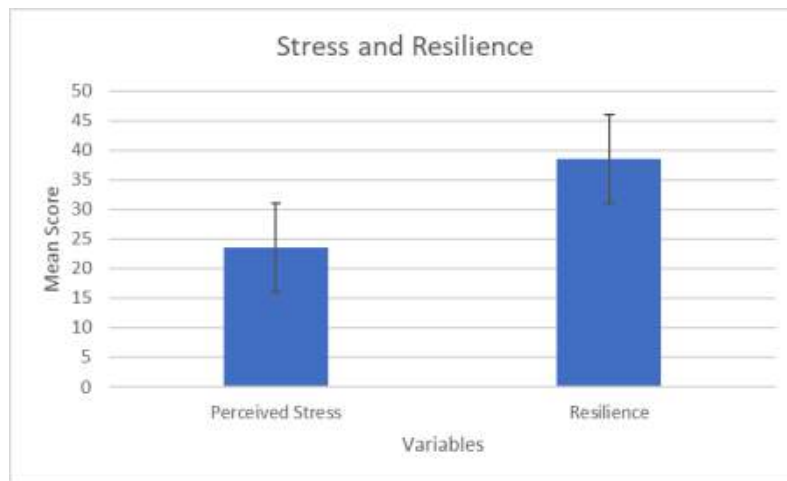


Fig 3: Relationship between Stress and Resilience among Indian Boarding School Youth (N = 25)

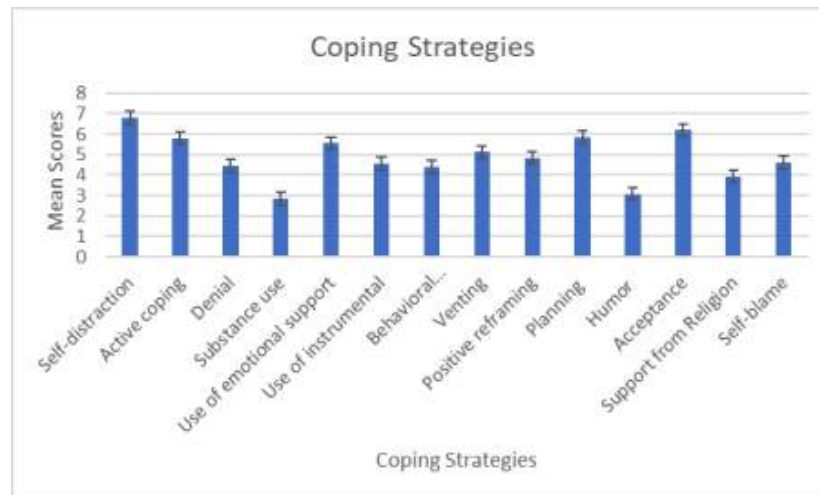


Fig 4: Coping Strategies among Indian Boarding School Youth (N = 25)

4. Discussion

The investigation sought to investigate the experiences of stress and coping among the students residing and studying in Indian boarding schools. Residential school students often encounter different types of mental and psychological challenges when away from family, as they are still navigating towards a more independent lifestyle. They face obstacles, such as academic challenges, social adjustments, and other demands of their daily routine. Accordingly the study is aimed at investigating these students' perceived stress, resilience, coping strategies, and emotion regulation. The research question for the study was:

“To explore the experiences of stress and coping strategies among Indian boarding school youth.” The study involved 25 participants, all of whom were currently students, enrolled in a boarding school in Bangalore, India. Participants were in the 10th or 11th grade, and their ages ranged between 15 and 17. These students came from diverse backgrounds. Some of them

lived in Bangalore or other cities in India, while the others lived in countries such as Saudi Arabia and Singapore. All the students shared similar routines and regular exposure to organised academic schedules and peer relationships, as well as limited regular communication with friends and relatives. Descriptive statistics were calculated to examine stress levels, resilience, emotion regulation, and coping strategies. The data on perceived stress indicated that most individuals encountered moderate levels of stress.

Similarly, the resilience scores were found to be moderate suggesting that students were able to adapt to different challenges despite being under stress. Coping strategies also indicated moderate use of coping strategies. Emotional regulation was also assessed using the Emotional Regulation Questionnaire. The results suggested high levels of cognitive reappraisal compared to the expressive suppression. This implies that instead of repressing their emotions, participants more often tried to understand and reinterpret feelings. The results validate the idea that students in boarding schools face noticeable stress but use coping mechanisms and mental strategies to manage and deal with these difficulties. Moderate levels of resilience, supported by the use of adaptive emotional regulation techniques, indicate that the students are able to adjust to the demands of residential schools and their lifestyle.

Overall, outcomes show that even though students attending boarding schools experience stress, they develop effective coping strategies that support their ability to deal with and maintain emotional well-being.

5. Conclusion

In conclusion, the findings of the investigation suggest that boarding school students experience moderate levels of stress while also demonstrating the effective use of coping methods and being able to maintain positive emotional well-being. Despite the challenges related to staying away from family, participants illustrated moderate levels of resilience, exhibiting the ability to adapt to tough academic and social circumstances. The use of coping methods was also moderate, showcasing a balanced approach to managing stress. Additionally, participants indicated that they utilised cognitive reappraisal more than expressive suppression, indicating a bias for adaptive emotion management approaches that include comprehending and interpreting emotions.

This data validates the idea that although boarding students face numerous challenges, they tend to develop effective coping methods that contribute to their emotional well-being. Overall, the investigation emphasises the importance of fostering and strengthening resilience and promoting the use of adaptive coping and emotional management skills to help students in residential schools.

Works Cited

- Connor-Davidson Resilience Scale (CD-RISC) Charles S. Carver (1997) *International Journal of Behavioral Medicine*, 4(1), 92–100. https://doi.org/10.1207/s15327558ijbm0401_6
- Deshpande, Apoorva, and Alka Ranjan. Resilience in Indian Adolescent Students: A Scoping Review. Vol. 13, 2025, ijip.in/wp-content/uploads/2025/12/18.01.174.20251304.pdf, <https://doi.org/10.25215/1304.174>. Accessed 20 Mar. 2026.
- Gross, J.J., & John, O.P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85, 348-362.
- John, et al. Self Report Measures for Love and Compassion Research: Personal Growth and Positive Emotions.
- Nandha, Mr, and Ms Nanthitha. “A Study on Stress Resilience and Coping among High School Students.” *International Journal of Humanities Social Science and Management (IJHSSM)*, vol. 4, no. 1, 2024, pp. 292–301, ijhssm.org/issue_dcp/A%20Study%20on%20Stress%20Resilience%20and%20Coping%20Among%20High%20School%20Students.pdf
- Kathryn M. Connor & Jonathan R.T. Davidson (2003) Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC). *Depression and Anxiety*, 18(2), 76–82. <https://doi.org/10.1002/da.10113>
- NovoPsych. “Coping Orientation to Problems Experienced Inventory (Brief-COPE).” NovoPsych, Mar. 2021, novopsych.com/assessments/formulation/brief-cope/.
- Redman-MacLaren, Michelle Louise, et al. “Measuring Resilience and Risk Factors for the Psychosocial Well-Being of Aboriginal and Torres Strait Islander Boarding School Students: Pilot Baseline Study Results.” *Frontiers in Education*, vol. 2, 7 Mar. 2017, <https://doi.org/10.3389/feduc.2017.00005>.
- Sheldon Cohen & Gail Williamsson, 1988, *Journal of Health and Social Behavior*, 24(4), 385–396. <https://doi.org/10.2307/2136404>
- Shrestha, Rajeev, et al. “Stress and Coping Mechanism among Students Residing in Private School Hostels.” *Mental Illness*, vol. 2023, 22 May 2023, p. e6535583, www.hindawi.com/journals/mij/2023/6535583/, <https://doi.org/10.1155/2023/6535583>.
- State of New Hampshire Employee Assistance Program. Perceived Stress Scale. 1983.

Computational Investigation of Structure-Property Relationships in Green Molecules Using Density Functional Theory: A Conceptual DFT Approach

By Yuven Joseph Amala Arasu

Abstract

Density Functional Theory (DFT) has emerged as a powerful computational tool for understanding molecular structure and reactivity while minimizing experimental resource use. In this study, DFT calculations were performed using Gaussian 09 to investigate a set of environmentally relevant molecules, including water, alcohols, acetic acid, and dihydroxybenzene isomers. Geometry optimization, frontier molecular orbital analysis, and conceptual DFT descriptors derived from Koopmans' theorem were used to evaluate molecular stability and reactivity. The results reveal clear distinctions between aliphatic and aromatic systems, as well as significant differences among structural isomers. Hydroquinone exhibited the highest reactivity, while water and simple alcohols showed greater stability. The findings highlight the importance of molecular structure in determining chemical behavior and demonstrate the utility of DFT in evaluating green molecules sustainably and efficiently.

Keywords Density Functional Theory, Koopmans' Theorem, Green Chemistry, B3LYP, Gaussian 09, HOMO-LUMO

Introduction

Density Functional Theory (DFT) is a quantum mechanical method used to investigate the electronic structure of atoms and molecules¹. Unlike wavefunction-based approaches, which depend on the coordinates of all electrons and require solving the Schrödinger equation with high computational cost, DFT focuses on the electron density as the fundamental variable. For a system with N electrons, wavefunction methods require $3N$ variables, whereas DFT depends only on three spatial variables, making it computationally efficient while retaining good accuracy. The theoretical foundation of DFT lies in the Hohenberg-Kohn theorems, which establish that the ground-state electron density uniquely determines all properties of a system². In practice, DFT calculations are performed using the Kohn-Sham formalism, where the interacting electron system is replaced by a system of non-interacting electrons in an effective potential³. Electron-electron interactions are accounted for through exchange-correlation functionals, which are approximated using methods such as Local Density Approximation (LDA), Generalized Gradient Approximation (GGA), and hybrid functionals like B3LYP. DFT enables the calculation of molecular geometries, total energies, dipole moments, and frontier molecular

¹ V. L. John, P. M. F., and T. P. V., "A DFT Study to Unravel the Fluorescence Mechanisms of APTES-Modified Carbon Dots," *Next Nanotechnology* 5 (2024), <https://doi.org/10.1016/j.nxnano.2023.100037>.

² P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Physical Review* 136, no. 3B (1964): B864–B871, <https://doi.org/10.1103/PhysRev.136.B864>.

³ Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Physical Review* 140, no. 4A (1965): A1133–A1138, <https://doi.org/10.1103/PhysRev.140.A1133>.

orbitals. The HOMO-LUMO gap is particularly important because it provides insight into chemical reactivity: smaller gaps indicate higher reactivity, whereas larger gaps suggest greater stability⁴. Additionally, DFT enables geometry optimization by minimizing atomic forces to determine the most stable molecular structure.

Green chemistry focuses on designing chemical processes and products that minimize the generation of hazardous substances⁵. It emphasizes prevention rather than remediation and is guided by principles such as waste minimization, safer solvents, energy efficiency, and the use of renewable resources. In this study, green molecules refer to small, environmentally benign compounds such as water, alcohols, and weak organic acids. Computational chemistry contributes to green chemistry by reducing the need for experimental trials, minimizing chemical waste, and enabling the prediction of molecular behavior before synthesis⁶. This study aims to investigate how molecular structure and solvation influence electronic properties and reactivity descriptors of selected green molecules using DFT.

DFT plays a central role in this study by providing a computational framework to evaluate the electronic structure and reactivity of molecules with a balance of accuracy and computational efficiency. By focusing on electron density rather than many-electron wavefunctions, DFT enables the calculation of key molecular properties such as optimized geometries, total energies, dipole moments, and frontier molecular orbitals. In this work, DFT is further extended through the use of conceptual descriptors derived from Koopmans' theorem, including ionization potential, electron affinity, chemical hardness, and electrophilicity index. These parameters allow for a deeper understanding of molecular stability, charge transfer tendencies, and reactivity patterns, which are essential for assessing the suitability of molecules within a green chemistry framework. The primary objective of this study is to investigate how molecular structure and substitution patterns influence electronic properties and reactivity in a set of environmentally relevant molecules. Specifically, the work aims to compare aliphatic compounds such as water, methanol, ethanol, and acetic acid with aromatic systems including catechol, resorcinol, and hydroquinone. By analyzing trends in frontier molecular orbitals and conceptual DFT descriptors, this study seeks to establish clear structure-property relationships that can inform the design and selection of green molecules. Ultimately, the work demonstrates how computational chemistry can be used as a predictive and sustainable tool to support the principles of green chemistry.

Computational Methodology

⁴ B. Ganiev, U. Mardonov, and G. Kholikova, "Molecular Structure, HOMO-LUMO, MEP Analysis of Triazine Compounds Using DFT (B3LYP) Calculations," *Materials Today: Proceedings* (2023), <https://doi.org/10.1016/j.matpr.2023.09.191>.

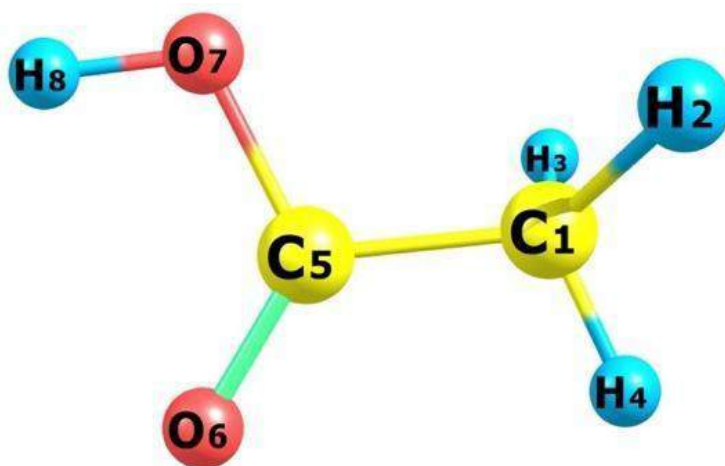
⁵ R. A. Sheldon, "Fundamentals of Green Chemistry: Efficiency in Reaction Design," *Chemical Society Reviews* 41, no. 4 (2012): 1437–1451, <https://doi.org/10.1039/C1CS15219J>.

⁶ L. Mammino, "Computational Chemistry and Green Chemistry: Familiarizing Chemistry Students with the Modes and Benefits of Promising Synergies," *Sustainable Chemistry and Pharmacy* 29 (2022), <https://doi.org/10.1016/j.scp.2022.100743>.

All calculations were performed using Gaussian 09⁷. Geometry optimizations were carried out using Density Functional Theory at the B3LYP/6-31G level of theory⁸. Frequency calculations were conducted to confirm that the optimized structures correspond to true minima, indicated by the absence of imaginary frequencies. Electronic properties, including HOMO and LUMO energies, dipole moments, and total energies, were extracted from the optimized structures. Ionization potential and electron affinity were estimated using Koopmans' theorem, and these values were used to derive conceptual DFT descriptors such as chemical hardness, chemical softness, chemical potential, electronegativity, and electrophilicity index⁹.

Results and Discussion

Acetic acid (CH_3COOH) consists of a methyl group attached to a carboxyl functional group. The calculated $\text{C}=\text{O}$ bond length (1.2328 Å) is significantly shorter than the $\text{C}-\text{O}$ single bond (1.3838 Å), reflecting the higher bond order of the carbonyl group. The $\text{O}-\text{H}$ bond length is 0.9823 Å, and the $\text{H}-\text{O}-\text{C}$ bond angle is 110.1° , consistent with a bent geometry around the oxygen atom. The $\text{C}-\text{C}$ bond length (1.4990 Å) confirms a single bond between an sp^3 methyl carbon and an sp^2 carbonyl carbon. Bond angles around the carbonyl carbon indicate trigonal planar geometry, while the methyl group retains tetrahedral geometry.



Bond	Bond length (Å)	Angle	Bond angle (°)
H8 - O7	0.98231	H8 - O7 - C5	110.073

⁷ M. J. Frisch et al., Gaussian 09, Revision A.1 (Wallingford, CT: Gaussian, Inc., 2009).

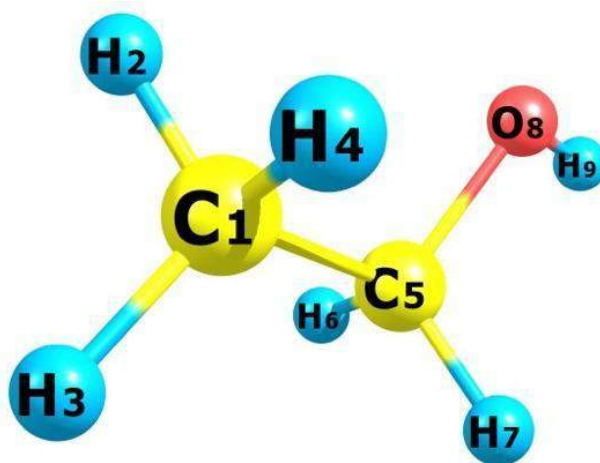
⁸ J. Tirado-Rives and W. L. Jorgensen, "Performance of B3LYP Density Functional Methods for a Large Set of Organic Molecules," *Journal of Chemical Theory and Computation* 4, no. 2 (2008): 297–306, <https://doi.org/10.1021/ct700248k>.

⁹ T. Koopmans, "Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms," *Physica* 1, no. 1–6 (1934): 104–113, [https://doi.org/10.1016/S0031-8914\(34\)90011-2](https://doi.org/10.1016/S0031-8914(34)90011-2).

Bond	Bond length (Å)	Angle	Bond angle (°)
O7 - C5	1.38380	O7 - C5 = O6	121.959
C5 = O6	1.23275	O6 - C5 - C1	126.719
C5 - C1	1.49898	C5 - C1 - H2	110.088
C1 - H4	1.08967	C5 - C1 - H3	110.087
C1 - H2 / C1 - H3	1.09477	C5 - C1 - H4	109.568

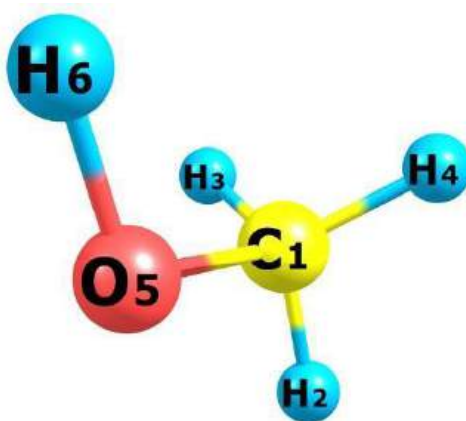
Fig 1: Optimised structure of Acetic acid, B3LYP/6-31G

Ethanol ($\text{C}_2\text{H}_5\text{OH}$) and methanol (CH_3OH) exhibit typical alcohol geometries. In ethanol, the C-O bond length (1.4596 Å) and O-H bond length (0.9779 Å) are consistent with expected values. The H-O-C angle (110.3°) reflects lone pair repulsion at the oxygen atom. The C-C bond length (1.5198 Å) confirms a single bond between sp^3 carbons. Methanol shows similar trends, with a C-O bond length of 1.4528 Å and an O-H bond length of 0.9778 Å. Minor variations in bond lengths and angles arise due to the electron-withdrawing effect of oxygen, but overall, the tetrahedral geometry is preserved around carbon.



Bond	Bond length (Å)	Angle	Bond angle (°)
C1 - H2 / C1 - H4 / C1 - H3	1.09347	H9 - O8 - C5	110.317
C1 - C5	1.51975	O8 - C5 - C1	106.892
C5 - H6 / C5 - H7	1.10058	H4 - C1 - C5 / H2 - C1 - C5	110.225
C5 - O8	1.45961	H3 - C1 - C5	110.652
O8 - H9	0.97787	H3 - C1 - H2 / H3 - C1 - H4	108.740
		H6 - C5 - H7	107.831
		H2 - C1 - H4	108.198
		O8 - C5 - H7 / O8 - C5 - H6	110.621
		H6 - C5 - C1 / H7 - C5 - C1	110.450

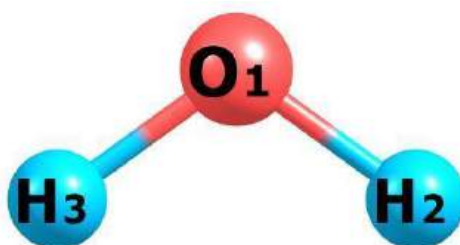
Fig 2: Optimised structure of Ethanol, B3LYP/6-31G



Bond	Bond length (Å)	Angle atoms	Bond angle (°)

H6 - O5	0.97777	H6 - O5 - C1	110.031
O5 - C1	1.45278	O5 - C1 - H2	105.792
C1 - H2	1.09060	O5 - C1 - H3 / O5 - C1 - H4	112.375
C1 - H3 / C1 - H4	1.09880	H2 - C1 - H3 / H2 - C1 - H4	108.613
		H3 - C1 - H4	108.912

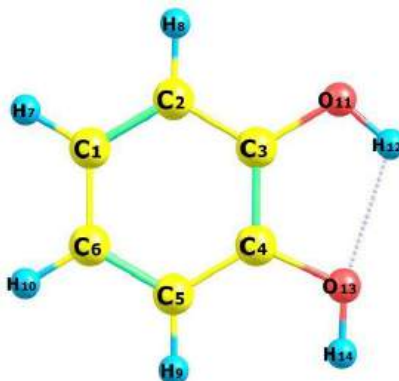
Fig 3: Optimised structure of Methanol, B3LYP/6-31G



Bond	Bond length (Å)	Angle	Bond angle (°)
H2 - O1 / H3 - O1	0.97592	H2 - O1 - H3	108.329

Fig 4: Optimised structure of Water, B3LYP/6-31G

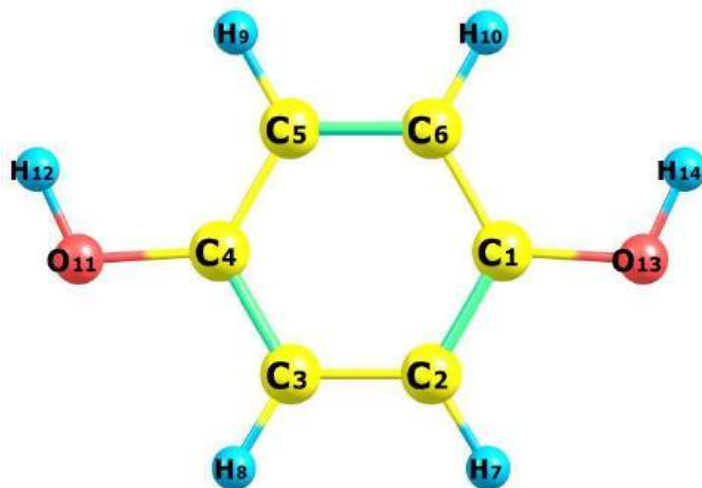
Water (H_2O) is a bent molecule due to the presence of two lone pairs on oxygen. The O-H bond length (0.9759 Å) and H-O-H angle (108.3°) are slightly smaller than the ideal tetrahedral angle, reflecting stronger repulsion from lone pairs. This geometry contributes to water's high polarity and strong hydrogen bonding capability.



Bond	Bond length (Å)	Bond (cont.)	Bond length (Å) (cont.)	Angle	Bond angle (°)	Angle (cont.)	Bond angle (°) (cont.)
H14 - O13	0.97402	C5 - H9	1.08688	H14 - O13 - C4	112.608	C1 - C6 - H10	120.424
O13 - C4	1.40255	C6 - C1	1.39894	H12 - O11 - C3	109.406	C6 - C1 - C2	120.388
H12 - O11	0.98000	C6 - H10	1.08449	O13 - C4 - C5	125.048	C6 - C1 - H7	120.083
O11 - C3	1.38563	C1 - C2	1.40114	O13 - C4 - C3	114.143	C2 - C1 - H7	119.530
C4 - C5	1.39185	C1 - H7	1.08465	C5 - C4 - C3	120.809	C1 - C2 - C3	119.728
C4 - C3	1.40444	C2 - C3	1.39302	C4 - C5 - C6	119.381	C1 - C2 - H8	121.754
C5 - C6	1.40305	C2 - H8	1.08343	C4 - C5 - H9	120.162	C3 - C2 - H8	118.518
				C6 - C5 - H9	120.458	C2 - C3 - C4	119.738
				C5 - C6 - C1	119.957	C2 - C3 - O11	120.100
				C5 - C6 - H10	119.619	C4 - C3 - O11	120.162

Fig 5: Optimised structure of Catechol, B3LYP/6-31G

Catechol (1,2-dihydroxybenzene) contains two hydroxyl groups in adjacent positions on an aromatic ring. The C–O bond lengths (~1.39-1.40 Å) are shorter than typical single bonds due to resonance interaction with the aromatic π -system. Intramolecular hydrogen bonding between the hydroxyl groups leads to slight distortions in bond angles and lengths. Resorcinol (1,3-dihydroxybenzene) contains hydroxyl groups in the meta position. The C-O bond lengths (1.3926 Å and 1.3914 Å) indicate partial delocalization, while the O-H bond lengths (~0.976 Å) and H-O-C angles (~111.6°) are consistent with hydroxyl group geometry. The aromatic C-C bond lengths confirm delocalization. Unlike catechol, resorcinol lacks intramolecular hydrogen bonding, resulting in a more uniform structure.

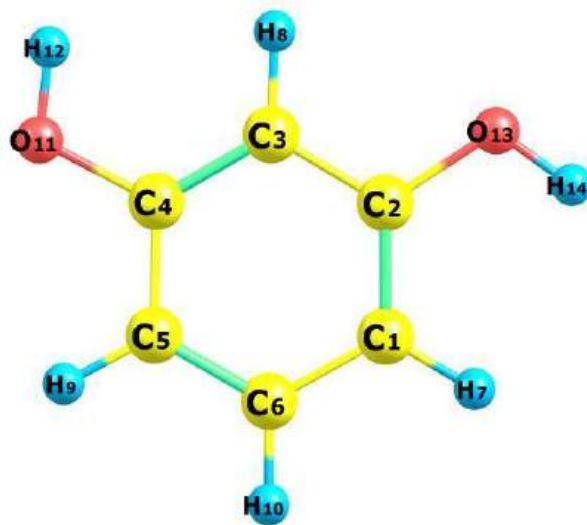


Bond	Bond length (Å)	Bond (cont.)	Bond length (Å) (cont.)	Angle	Bond angle (°)	Angle (cont.)	Bond angle (°) (cont.)
H14 - O13	0.97595	C2 - H7	1.08331	H14 - O13 - C1	111.62484	C4 - C3 - H8	118.77271
O13 - C1	1.39662	C3 - C4	1.40133	H12 - O11 - C4	111.62484	O11 - C4 - C3	116.78875
H12 - O11	0.97595	C3 - H8	1.08331	O13 - C1 - C2	116.78875	O11 - C4 - C5	123.10988
O11 - C4	1.39662	C4 - C5	1.39700	O13 - C1 - C6	123.10988	C3 - C4 - C5	120.10137
C1 - C2	1.40133	C5 - C6	1.40156	C2 - C1 - C6	120.10137	C4 - C5 - C6	119.89112
C1 - C6	1.39700	C5 - H9	1.08676	C1 - C2 - C3	120.00751	C4 - C5 - H9	120.49583
C2 - C3	1.39231	C6 - H10	1.08676	C1 - C2 - H7	118.77271	C6 - C5 - H9	119.61306
				C3 - C2 - H7	121.21978	C5 - C6 - C1	119.89112
				C2 - C3 - C4	120.00751	C5 - C6 - H10	119.61305
				C2 - C3 - H8	121.21978	C1 - C6 - H10	120.49583

Fig 6: Optimised structure of Hydroquinone, B3LYP/6-31G

Hydroquinone (1,4-dihydroxybenzene) exhibits para-substitution, resulting in a highly symmetrical structure. Equivalent bond lengths and angles reflect this symmetry, and the

aromatic ring shows uniform delocalization. Compared to the other isomers, hydroquinone displays the least structural distortion.



Bond	Bond length (Å)	Bond (cont.)	Bond length (Å) (cont.)	Angle atoms	Bond angle (°)	Angle atoms (cont.)	Bond angle (°) (cont.)
H14 - O13	0.97609	C2 - C3	1.39881	H14 - O13 - C2	111.57069	C4 - C3 - H8	122.06538
O13 - C2	1.39257	C3 - C4	1.39635	H12 - O11 - C4	111.75349	O11 - C4 - C3	122.21059
H12 - O11	0.97630	C3 - H8	1.08475	O13 - C2 - C1	122.83732	O11 - C4 - C5	116.73846
O11 - C4	1.39136	C4 - C5	1.40190	O13 - C2 - C3	116.20355	C3 - C4 - C5	121.05095
C1 - C2	1.39900	C5 - C6	1.39448	C1 - C2 - C3	120.95914	C4 - C5 - C6	118.89388
C1 - C6	1.40151	C5 - H9	1.08256	C2 - C1 - C6	118.90006	C4 - C5 - H9	119.10183
C1 - H7	1.08610	C6 - H10	1.08509	C2 - C1 - H7	120.57067	C6 - C5 - H9	122.00428
				C6 - C1 - H7	120.52926	C1 - C6 - C5	121.13918
				C2 - C3 - C4	119.05678	C1 - C6 - H10	119.32922
				C2 - C3 - H8	118.87783	C5 - C6 - H10	119.53160

Fig 7: Optimised structure of Resorcinol, B3LYP/6-31G

Analysis of frontier molecular orbitals and conceptual DFT descriptors reveals clear trends across the dataset. The HOMO and LUMO orbitals provide insight into the electron distribution and reactivity of the molecules. Aromatic systems exhibit more delocalized HOMO orbitals, while aliphatic molecules show localized electron density. The HOMO-LUMO gap further supports trends in reactivity, with smaller gaps corresponding to increased chemical reactivity. The reduced HOMO-LUMO gap in aromatic systems implies their enhanced reactivity compared to aliphatic molecules.

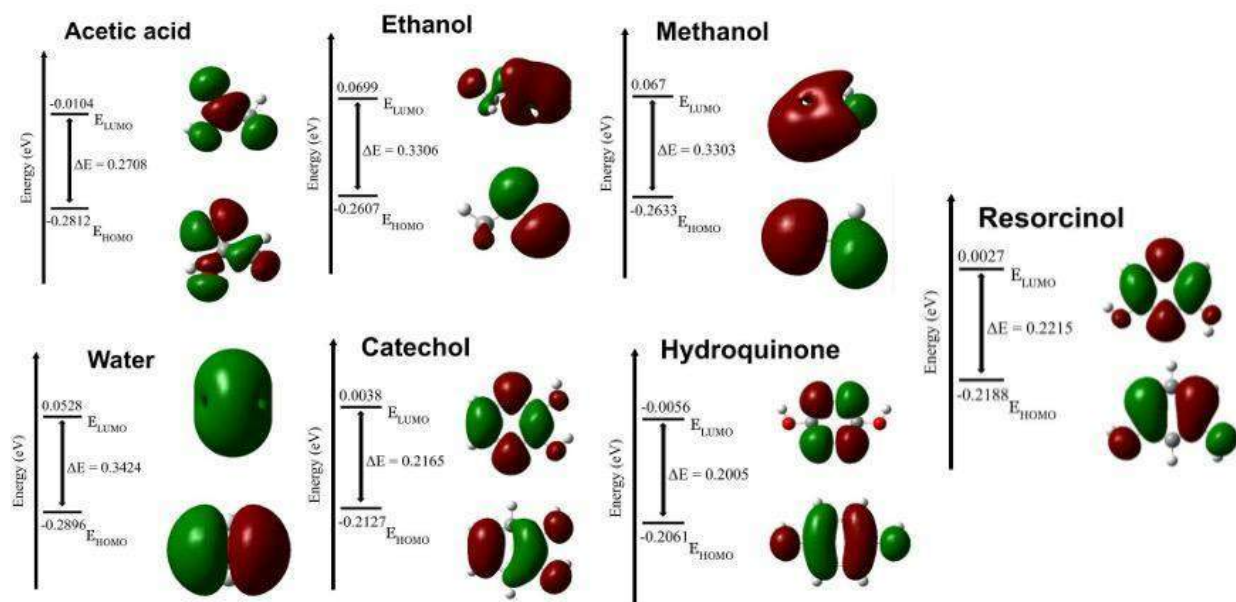


Fig 8: The HOMO-LUMO diagrams of the molecules under study

Aromatic molecules such as catechol, resorcinol, and hydroquinone exhibit lower ionization potentials and lower chemical hardness compared to aliphatic molecules, indicating higher reactivity. This behavior arises from the presence of delocalized π -electrons, which facilitate electron transfer. Among the aromatic compounds, hydroquinone shows the lowest hardness and highest electrophilicity, indicating the greatest reactivity. Catechol and resorcinol exhibit slightly higher hardness values, with catechol influenced by intramolecular hydrogen bonding and resorcinol by reduced resonance interaction in the meta configuration. In contrast, water, methanol, and ethanol display higher hardness and lower electrophilicity, indicating greater stability and lower reactivity. These properties align with their role as stable, environmentally benign solvents.

System	Ionization Potential (I)	Electron Affinity (A)	Chemical Hardness (η)	Chemical softness (S)	Chemical potential (μ)	Electronegativity (χ)	Electrophilicity Index (ω)
Acetic acid	0.2812	0.0104	0.1354	3.6928	-0.1458	0.1458	0.0785
Ethanol	0.2607	-0.0699	0.1653	3.0248	-0.0954	0.0954	0.0275
Methanol	2633	-0.067	0.16515	3.0276	-0.0982	0.0982	0.0292
Water	0.2896	-0.0528	0.1712	2.9206	-0.1184	0.1184	0.0409
Catechol	0.2127	-0.0038	0.10825	4.6189	-0.1045	0.1045	0.0504
Hydroquinone	0.2061	0.0056	0.10025	4.9875	-0.1059	0.1059	0.0559
Resorcinol	0.2188	-0.0027	0.11075	4.5147	-0.1081	0.1081	0.0528

Table 1. Conceptual DFT descriptors derived from Koopman's Theorem

A key finding of this study is the influence of substitution patterns on electronic properties. Among the dihydroxybenzene isomers, hydroquinone (para) exhibits the highest degree of electron delocalization due to its symmetrical structure. Catechol (ortho) shows deviations due to intramolecular hydrogen bonding, while resorcinol (meta) displays intermediate behavior due to limited resonance interaction between substituents. These results demonstrate that even small changes in molecular structure can significantly influence electronic properties and reactivity.

The electronic properties of the selected molecules were evaluated using frontier molecular orbital theory and conceptual DFT descriptors derived through Koopmans' theorem. Parameters such as ionization potential, electron affinity, chemical hardness, softness, chemical potential, electronegativity, and electrophilicity index were used to assess molecular stability and reactivity. These descriptors provide a quantitative framework to compare charge transfer tendencies and overall chemical behavior across both aliphatic and aromatic systems.

Clear trends are observed when comparing the molecules. Water, methanol, and ethanol exhibit higher chemical hardness values, indicating greater stability and lower reactivity. This is consistent with their saturated structures and lack of conjugation. In contrast, the aromatic dihydroxybenzene derivatives show significantly lower hardness values, suggesting increased reactivity. This behavior arises from the presence of delocalized π electrons in the aromatic ring, which facilitates charge transfer and stabilizes transient electronic states. The electrophilicity index further highlights differences in reactivity. Hydroquinone displays the highest electrophilicity, followed by resorcinol and catechol, indicating a stronger tendency to participate in charge transfer interactions. Alcohols such as ethanol and methanol show much lower electrophilicity, reinforcing their relatively inert nature. Ionization potential values support these

observations, with aromatic compounds requiring less energy to remove an electron compared to water and acetic acid, reflecting enhanced electron delocalization. Hydroquinone, with para substitution, shows the lowest hardness and highest electrophilicity, indicating greater electron delocalization. Catechol exhibits slightly higher hardness, likely due to intramolecular hydrogen bonding, while resorcinol shows intermediate behavior because the meta substitution limits resonance interaction. These results emphasize that even subtle changes in substitution pattern can lead to meaningful differences in electronic structure and reactivity. Negative electron affinity values observed for alcohols and water indicate their low tendency to accept electrons, consistent with their saturated electronic structure.

From a green chemistry perspective, molecules with high stability, moderate reactivity, and strong polarity are desirable. Water and alcohols exhibit these properties, making them ideal green solvents. Phenolic compounds, while more reactive, play important roles in redox chemistry and may contribute to environmentally relevant processes such as antioxidant activity. DFT provides a powerful tool for evaluating these properties without the need for extensive experimental work, aligning with the principles of green chemistry by minimizing waste and resource use. This study is limited by the small number of molecules. Additionally, Koopmans' theorem provides only an approximation of ionization potential and electron affinity, and more accurate methods could be employed in future studies.

Conclusions

This study demonstrates the effectiveness of DFT as a robust and efficient approach for investigating the structural and electronic properties of environmentally relevant molecules. By employing the B3LYP/6-31G level of theory within Gaussian 09, reliable optimized geometries and electronic descriptors were obtained for a diverse set of small molecules, including aliphatic compounds and aromatic dihydroxybenzene isomers. The results confirm that DFT provides meaningful insights into molecular stability and reactivity without the need for extensive experimental procedures, thereby aligning with the principles of sustainable and resource-efficient research.

The structural analysis revealed that the optimized geometries closely follow expected chemical principles, with variations in bond lengths and angles reflecting differences in hybridization, functional groups, and electronic environments. Aliphatic molecules such as water, methanol, ethanol, and acetic acid exhibited localized bonding and relatively uniform geometries, while aromatic systems displayed clear evidence of electron delocalization within the benzene ring. Subtle structural differences among the dihydroxybenzene isomers were also identified, particularly in relation to hydrogen bonding and substitution patterns, which were found to influence the overall symmetry and electronic distribution of the molecules. A key outcome of this work is the identification of clear trends in electronic properties across the dataset. Conceptual DFT descriptors derived from Koopmans' theorem revealed that aliphatic molecules possess higher chemical hardness and lower electrophilicity, indicating greater stability and lower reactivity. In contrast, aromatic compounds exhibited reduced hardness and

increased electrophilicity, reflecting enhanced charge transfer capability due to π -electron delocalization. Among the aromatic systems, hydroquinone demonstrated the lowest hardness and highest electrophilicity, highlighting its greater reactivity compared to catechol and resorcinol. These differences were directly linked to substitution patterns, where para substitution promotes more effective electron delocalization than ortho or meta configurations.

The study further emphasizes the importance of structure–property relationships in determining molecular behavior. Even small changes in the position of functional groups were shown to produce measurable differences in electronic properties, underscoring the sensitivity of molecular systems to structural variation. This observation is particularly significant for the rational design of molecules, where tuning substitution patterns can be used to modulate reactivity and stability for specific applications. From a green chemistry perspective, the findings reinforce the suitability of water and simple alcohols as stable, low-reactivity solvents, while also highlighting the functional importance of more reactive phenolic compounds in redox and antioxidant processes. The ability to predict such behavior computationally reduces reliance on trial-and-error experimentation, thereby minimizing chemical waste, energy consumption, and potential environmental impact. Overall, this work highlights the value of computational chemistry as a powerful and sustainable tool for understanding and predicting molecular properties. By establishing clear connections between molecular structure and electronic behavior, the study provides a foundation for the informed selection and design of green molecules, contributing to the broader goal of developing environmentally responsible chemical processes.

Works Cited

- John, V. L., P. M. F., and T. P. V. "A DFT Study to Unravel the Fluorescence Mechanisms of APTES-Modified Carbon Dots." *Next Nanotechnology* 5 (2024). <https://doi.org/10.1016/j.nxnano.2023.100037>.
- Hohenberg, P., and W. Kohn. "Inhomogeneous Electron Gas." *Physical Review* 136, no. 3B (1964): B864–B871. <https://doi.org/10.1103/PhysRev.136.B864>.
- Kohn, W., and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects." *Physical Review* 140, no. 4A (1965): A1133–A1138. <https://doi.org/10.1103/PhysRev.140.A1133>.
- Ganiev, B., U. Mardonov, and G. Kholikova. "Molecular Structure, HOMO-LUMO, MEP Analysis of Triazine Compounds Using DFT (B3LYP) Calculations." *Materials Today: Proceedings* (2023). <https://doi.org/10.1016/j.matpr.2023.09.191>.
- Sheldon, R. A. "Fundamentals of Green Chemistry: Efficiency in Reaction Design." *Chemical Society Reviews* 41, no. 4 (2012): 1437–1451. <https://doi.org/10.1039/C1CS15219J>.
- Mammino, L. "Computational Chemistry and Green Chemistry: Familiarizing Chemistry Students with the Modes and Benefits of Promising Synergies." *Sustainable Chemistry and Pharmacy* 29 (2022). <https://doi.org/10.1016/j.scp.2022.100743>.
- Frisch, M. J., et al. *Gaussian 09, Revision A.1*. Wallingford, CT: Gaussian, Inc., 2009.
- Tirado-Rives, J., and W. L. Jorgensen. "Performance of B3LYP Density Functional Methods for a Large Set of Organic Molecules." *Journal of Chemical Theory and Computation* 4, no. 2 (2008): 297–306. <https://doi.org/10.1021/ct700248k>.
- Koopmans, T. "Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms." *Physica* 1, no. 1–6 (1934): 104–113. [https://doi.org/10.1016/S0031-8914\(34\)90011-2](https://doi.org/10.1016/S0031-8914(34)90011-2).

The Cardiovascular Impact of Wildfire Smoke in Today's Changing Climate

By Jessica Lipshutz

The physical devastation from the Altadena and Palisades wildfires in January 2025 continues to be noticeable to Los Angeles residents - empty lots, blackened trees, a favorite restaurant no longer visited. However, what may not be considered is the long term effects of wildfire particulate matter on the health of residents in Los Angeles and throughout the United States. Global warming has intensified wildfire seasons, and experts foresee an expansion of both the number of days and areas affected by wildfires. And while the harmful effects of wildfire smoke on our lungs is well-recognized, its impact on cardiovascular health remains largely underestimated. In fact, cardiovascular disease is the fourth leading health consequence, ahead of respiratory diseases and cancers (Sagheer et al.). And although elderly individuals and those with pre-existing heart conditions are the most vulnerable, the risk applies to people of all ages, even the young and healthy. As wildfire seasons grow longer and more severe due to climate change, the cardiovascular consequences of smoke exposure represent an urgent, underestimated public health crisis.

Wildfire smoke contains fine particulate matter (PM). The most toxic subtype of PM is PM_{2.5}, which is defined by its size of 2.5 micrometers or less. Its small size allows PM_{2.5} to penetrate deep into the lungs and enter the bloodstream, triggering widespread inflammation and a whole host of other harmful effects on our body including increased risk of blood clotting and “oxidative stress” which in lay terms means harmful versions of oxygen species that can cause injury to our cells. When blood vessel linings become damaged, a condition known as endothelial dysfunction occurs, which increases the risk of heart attacks, arrhythmias, and strokes. Thus, the heart and cardiovascular system are uniquely vulnerable to wildfire smoke exposure.

The link between PM_{2.5} from wildfire smoke and excess cardiovascular risk is laid out clearly in an article published by Stacey Alexeeff, PhD and others at Kaiser Permanente Northern California in *Journal of the American Heart Association* last year. The researchers studied air quality during the second half of 2018, when two large wildfires burned in Northern California - the Mendocino Complex Wildfire (July - August) and the Camp Fire (November). High PM_{2.5} concentrations during the Mendocino Complex Fire were associated with a 23.1% increase in cardiovascular events (which included heart attack, stroke, and heart failure) and a 35.8% rise in mortality compared to when PM_{2.5} levels were in a “good” range. Interestingly, the 2018 Camp Fire was not associated with an increase in heart-related events, likely due to heightened public awareness after the Mendocino fire resulting in school closures and more protective mask use; the study documented more Google searches for “N95” during the Camp Fire time period compared to previous fires. Another theory is termed the “depletion of susceptibles” —those most vulnerable had already experienced events during exposure a few months earlier during the Mendocino fires (Beck). These findings illustrate the profound effect of people being aware and educated about the health risks of wildfire smoke.

In 2024, Dr. Kai Chen at the Yale School of Public Health and colleagues published a study in *Proceedings of the National Academy of Sciences* that looked at average 12-month PM_{2.5} exposure across the US between 2007 and 2020, and found higher PM_{2.5} levels were significantly associated with mortality not only from cardiovascular disease but also diabetes, kidney disease, and mental illness. Researchers estimate that 4,512 excess cardiovascular deaths occur annually in the U.S. due to wildfire smoke, with elevated risks even very apparent at low concentrations, such as 0.1-0.2ug/m³ which is well within an EPA-defined “Good” range. Notably, this risk from wildfire PM_{2.5} was even found outside of known wildfire zones, highlighting how this issue is relevant to all Americans (Chen et al.).

Wildfires often coincide with other climate events, such as heatwaves, creating a deadly synergy, as described by Dr. Chen in an editorial in the *Journal of the American College of Cardiology (JACC)*. Combined exposure to wildfire smoke and extreme heat raises cardiac death risk by 86%, surpassing the individual effects of heat or smoke alone (Chen et al.). Beyond smoke inhalation, wildfires worsen health risks through psychological stress, reduced physical activity, disrupted sleep, and dehydration, all compounding cardiovascular strain. In the face of climate change, the heat - smoke combination is likely to become more frequent, accentuating the crucial need for comprehensive preparation of these events.

Studies of firefighters demonstrate acute symptoms from wildfire smoke exposure including chest pain, shortness of breath, dizziness, and elevated heart rates. The unique toxicity of wildfire smoke—burning plastics, vehicles, and homes—introduces heavy metals and carcinogens, intensifying cardiovascular strain. Chronic, repeated exposure has led firefighters in their 30s to develop early signs of cardiovascular disease, including pulmonary emboli and chest pain. Cultural pressures to avoid mask use and downplay symptoms further perpetuate very preventable harm (Dreier).

Monitoring air quality indices and limiting exposure to fine matter particulates are critical, especially for those with cardiovascular conditions. Public awareness—evident in the increased awareness to wear N95 masks during the Camp Fire—may help reduce cardiovascular events. Beyond masks that filter PM_{2.5}, air purifiers might be effective at lowering cardiovascular risk associated with wildfire smoke PM_{2.5} exposure. An August 6, 2025 *JACC* study demonstrated HEPA filtration lowers indoor PM_{2.5} and reduces systolic blood pressure by 3.0 mmHg in individuals with elevated baseline levels, corresponding to a 4–6% lower risk of major adverse cardiac events (Brugge et al.). Wildfire smoke is a pressing public health concern with severe health consequences, particularly when it comes to our hearts. This risk affects individuals living in fire zones like Los Angeles County, but also affects people throughout the United States. Climate change has advanced California’s fire season by six weeks since 1990, and extreme events are becoming more frequent and severe. Studies predict that wildfire smoke could claim 70,000 U.S. lives annually by 2050. Despite its importance, federal research funding in this area has been curtailed; however, the evidence clearly shows us that more data collection and analysis, not less, is needed on this topic. We need greater attention, and stricter controls, on

national health air quality standards. The varying results of the Mendocino and Camp Fires highlight the critical role of awareness and preparedness. Expanding research and implementing interventions such as air purifiers, masks, adequate ventilation, and public education are essential to protect heart health in an era of intensifying wildfire threats.

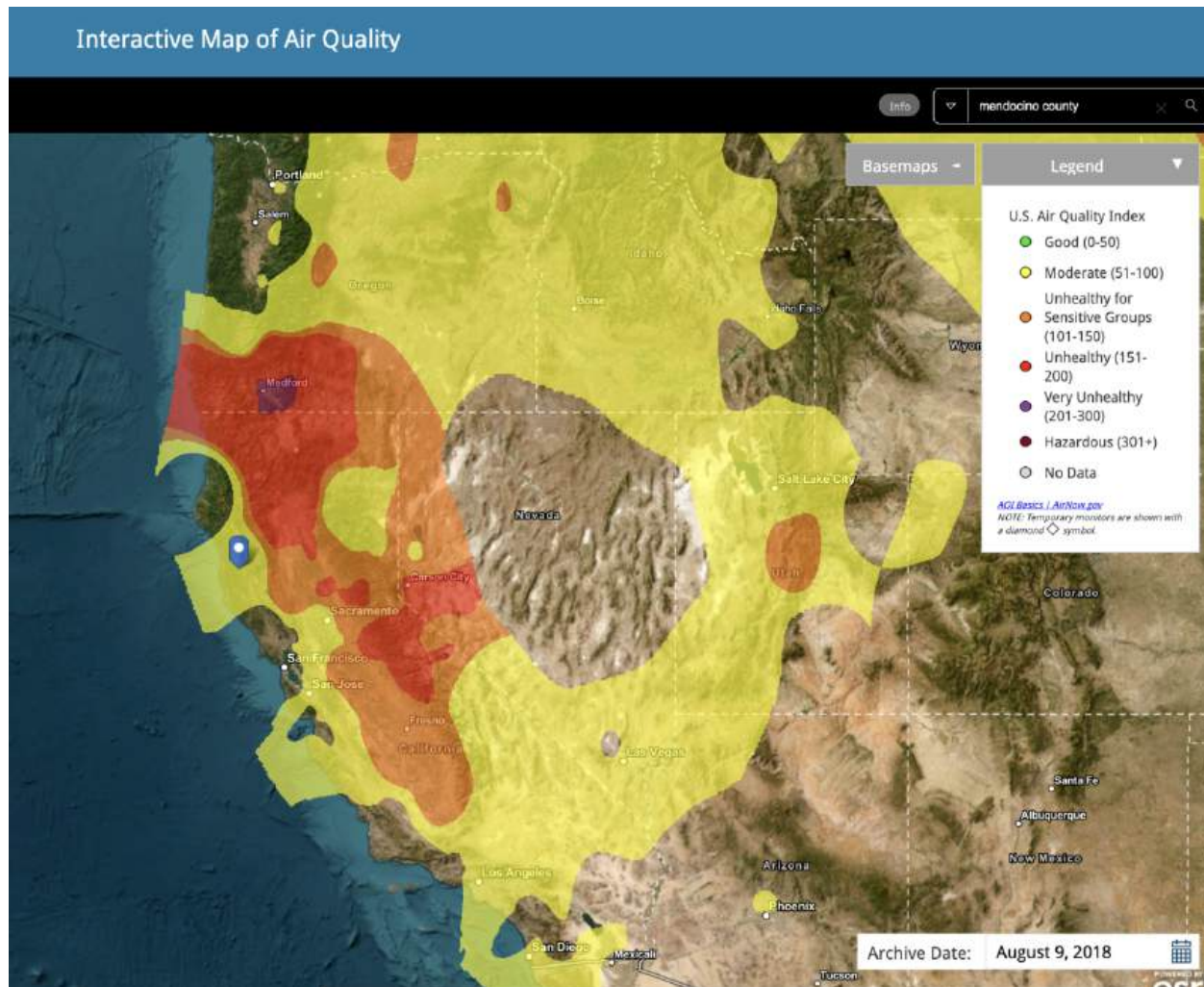


Image 1: Air Quality, based on PM_{2.5} concentration, in the area of the Mendocino Complex Fire in 2018. From AirNow.gov

Work Cited

- Beck, Debra L. "Cover Story: A New Threat to Cardiovascular Health: Climate Change, Extreme Weather and Wildfires." *American College of Cardiology*, 3 Apr. 2025.
<https://www.acc.org/latest-in-cardiology/articles/2025/04/01/01/cover-story-a-new-threat-to-cardiovascular-health>
- Sagheer, Usman, et al. "Environmental pollution and cardiovascular disease." *JACC: Advances*, vol. 3, no. 2, Feb. 2024, p. 100805.
<https://doi.org/10.1016/j.jacadv.2023.100805>
- Alexeeff, Stacey E., et al. "Wildfire air pollution and rates of cardiovascular events and mortality in Northern California in 2018." *Journal of the American Heart Association*, vol. 14, no. 4, 18 Feb. 2025.
<https://doi.org/10.1161/jaha.124.036264>
- Chen, Kai, et al. "Wildfires, compound extreme events, climate change, and Cardiovascular Health." *Journal of the American College of Cardiology*, vol. 85, no. 12, Apr. 2025, pp. 1379–1381
<https://doi.org/10.1016/j.jacc.2025.01.009>
- Mulkey, Sachi, and Harry Stevens. "Wildfire Smoke Will Kill Thousands More by 2050, Study Finds." *New York Times*.
www.nytimes.com/2025/09/18/climate/wildfire-smoke-health-deaths.html. Accessed 22 Apr. 2026
- Brugge, Doug, et al. "Effect of HEPA filtration air purifiers on blood pressure." *JACC*, vol. 86, no. 8, Aug. 2025, pp. 577–589.
<https://doi.org/10.1016/j.jacc.2025.06.037>
- Dreier, Hannah. "Wildfire Fighters, Unmasked in Toxic Smoke, Are Getting Sick and Dying - The New York Times." *New York Times*.
www.nytimes.com/2025/08/17/us/wildfire-firefighters-masks-smoke.html. Accessed 22 Apr. 2026.

Digital Allowance Access, Buying Autonomy, and Impulsive Buying among Adolescents: An Exploratory Study By Suhana Shah

Abstract

The present study focused on the correlations between digital allowance availability, buying autonomy, and impulsive buying behavior among adolescents. A total of 25 participants aged between 14 and 17 years participated in the online survey and responded to Likert scale questions. The correlation analysis demonstrated a low positive correlation between digital allowance availability and buying autonomy ($r = .13$), as well as a low positive correlation between digital allowance availability and impulsive buying behavior ($r = .16$). In addition, a moderate negative correlation was found between buying autonomy and impulsive buying behavior ($r = -.31$). At the same time, all these relationships were insignificant due to the small number of participants included in the analysis. Therefore, it can be concluded that buying autonomy might protect teenagers from impulsive purchases. Nevertheless, it is not possible to draw any final conclusions regarding the problem under consideration because more research efforts are needed to address the existing knowledge gap.

Keywords digital allowance, buying autonomy, impulsive buying behavior, adolescents, financial decision-making, consumer behavior, self-regulation, digital payments, youth consumption, exploratory study

1. Introduction

This period is characterized by significant social and psychological changes since adolescents start to develop their identities and become more independent from other people. With age, adolescents gain increased freedom when making different types of choices since they do not entirely rely on their parents or caregivers. They start taking control over some aspects of their lives independently from other people, including financial management. The freedom of managing finances and purchasing goods is crucial in learning how to act responsibly in adulthood.

Allowances are a significant tool for teenagers when learning how to handle money correctly. The fact that adolescents get allowances gives them an opportunity to deal with personal finances independently, which helps them gain valuable experience about managing their money and purchases. Thus, allowing adolescents to control their allowance money contributes to their personal financial responsibility (Rocha et al. 2023).

Currently, adolescents have fewer opportunities to receive money in cash since technology has changed the way money is managed today. The possibility of getting money using bank transactions and other tools made it possible to introduce the concept of allowance digital access to the attention of teenagers. Digital technology allows adolescents to receive allowance money electronically without cash transactions. Digital money can change the way people think about spending. Unlike cash, digital payments are fast and convenient, and they do

not require physically handing over money. Because of this, spending digitally may feel easier and less restrictive. The convenience of digital payments may increase adolescents' sense of independence, but it may also make spending more frequent or less carefully considered. (Bhakat and Muruganantham; Bellini and Aiolfi).

Autonomy refers to the ability to make decisions independently. During adolescence, gaining autonomy is an important part of development because it allows young people to practice making choices and taking responsibility for those choices. "Buying autonomy" specifically refers to the freedom adolescents have when deciding how to spend their money. As adolescents gain more control over their financial resources, they often begin to make more of their own purchasing decisions without direct parental involvement. (Li et al.).

This independence can have both positive and negative effects. On one hand, having control over spending can help adolescents develop responsibility and improve their decision-making skills. On the other hand, increased independence may also lead to riskier choices if adolescents make decisions without enough experience or guidance.

Impulsive buying refers to purchases that are made suddenly and without much planning. These purchases are often driven by emotions, excitement, or the desire for immediate satisfaction rather than careful thinking about the consequences. (Rook and Fisher) Adolescents may be more likely to engage in impulsive buying because their self-control and decision-making abilities are still developing. Emotional reactions and the appeal of immediate rewards can strongly influence their choices. As a result, adolescents may sometimes make quick purchasing decisions without fully considering whether they really need the item. (Lin, 2005)

Although impulsive buying has been widely studied among adult consumers, fewer studies have focused specifically on adolescents. This is important because many financial habits and spending behaviors begin to form during adolescence. While research on financial behavior continues to grow, several gaps still exist. There is limited research examining how digital allowance access affects adolescents' spending behaviors. In addition, relatively few studies have focused on impulsive buying among adolescents, especially in relation to their level of independence in financial decision-making. Furthermore, little research has explored the relationship between digital allowance access, buying autonomy, and impulsive buying behavior together. Understanding how these factors interact may provide valuable insights into how adolescents make financial decisions in a digital world. The present study aims to explore the relationship between digital allowance access, buying autonomy, and impulsive buying behavior among adolescents. This research takes an exploratory approach to examine whether access to digital allowances is related to adolescents' level of buying autonomy and their tendency to make impulsive purchases. By examining these relationships, the study seeks to contribute to a better understanding of how digital financial systems may influence adolescent spending behavior. (Nyrhinen et al.; Bhardwaj et al.)

1.1 Objectives of the Study

1. To assess the extent of digital allowance access among adolescents.

2. To examine the association between digital allowance access and buying autonomy among adolescents.
3. To examine the association between buying autonomy and impulsive buying behavior among adolescents.
4. To examine the association between digital allowance access and impulsive buying behavior among adolescents.

2. Research Methodology

2.1 Participants

The sample consisted of 25 adolescents aged between 14 and 17 years. Of the total participants, 72% were female and 28% were male. With regard to educational background, 96% of the participants were enrolled in private institutions, while 4% attended a government school.

2.2 Measures

a. Digital Allowance Access

Digital allowance access was measured using a seven-item self-developed scale designed to assess adolescents' access to and independence in using digital payment methods (e.g., UPI, e-wallets). The questions tested the availability of digital payment options, frequency of money receipt through digital payment methods, confidence with digital payment apps, and perception of independence while spending digital allowance. For instance, "I am able to use digital money at any time when I need something to buy" and "I don't have to ask anyone's permission when making digital payments". Answers were provided based on the five-point Likert scale where participants could choose between Strongly Disagree (1 point) and Strongly Agree (5 points). High scores meant increased digital allowance availability. Due to being little researched among adolescents, context-specific questionnaire items were created.

b. Buying Autonomy

Autonomous purchase behavior was assessed through five items derived from the Autonomous Decision Making facet of the Adolescent Autonomy Questionnaire (AAQ; Li et al., 2022). This questionnaire evaluates adolescents' self-perceived freedom and self-reliance while making decisions. In relation to this, an example item is "I am able to pay digitally when I want to purchase something." Another is "I make my decision about my digital allowance expenditures independently." In this research, these items were revised based on purchase behaviors to align with the current research purpose. Items were evaluated using a five-point Likert scale ranging from 1=Strongly Disagree to 5=Strongly Agree.

c. Impulsive Buying Behavior

The impulse purchasing behavior scale comprised nine items adopted from the Impulsive Buying Scale created by Rook and Fisher (1995). This scale gauges an individual's predisposition to indulge in impulsive shopping behavior spurred by sudden desires rather than well-reasoned decisions. For example, "I often buy things without thinking carefully" and "When I see something I like, I feel like buying it immediately". Minor wording adjustments were made to ensure age-appropriate language for adolescents. Responses were recorded on a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). One item assessing planned purchasing was reverse-coded prior to analysis. Higher scores indicated greater impulsive buying tendencies.

2.3 Procedure

Data were collected using an online survey administered through Google Forms. A convenience sampling method was employed to recruit participants. The survey link was distributed to eligible adolescents, who completed the questionnaire online at their convenience.

2.4 Ethical Considerations

As all participants were below 18 years of age, informed consent was obtained from both parents and the adolescents prior to participation. Participation was voluntary, and participants were informed of their right to withdraw at any time without penalty. Confidentiality and anonymity were ensured, and no personally identifying information was collected.

2.5 Data Analysis

Data was analyzed using Microsoft Excel. Descriptive statistics, including means and standard deviations, were computed to summarize the data. Pearson's correlation analysis was conducted to examine the relationships among the study variables.

3. Results

The sample consisted of adolescents from Grades 9 to 12, with participants drawn from Dubai and Belgium. Monthly allowance amounts showed substantial variation, ranging from 0 AED to 2,500 AED. Allowances were given through both cash and digital modes (see Fig. 1), 40% of participants received a combination of cash and digital payments, 32% received cash only, and 28% received allowances exclusively through digital means.

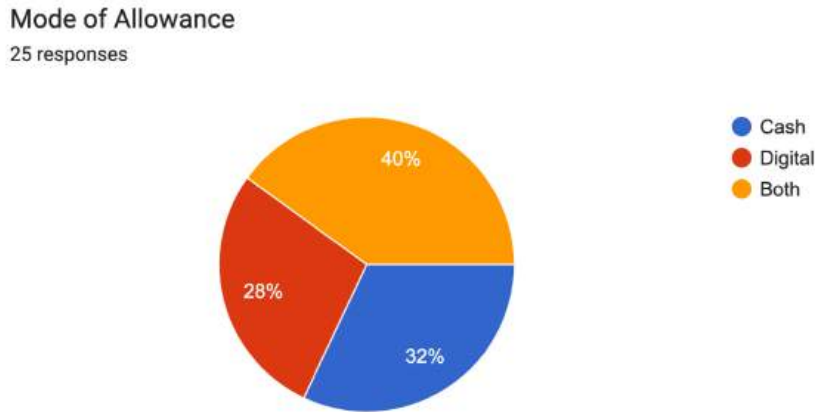


Figure 1. Mode of Allowance

Descriptive statistics were computed for three variables (Refer Table 1) namely digital allowance access, impulsive buying behavior, and buying autonomy. Buying autonomy had the highest mean score ($M = 4.40$, $SD = 0.52$), indicating generally high perceived independence in purchasing decisions along with low variability among participants. Impulsive buying behavior demonstrated a moderate mean ($M = 3.22$, $SD = 0.83$), suggesting moderate levels of spontaneous purchasing tendencies with some variability. Digital allowance access showed the lowest mean ($M = 2.00$, $SD = 1.06$) and the largest standard deviation, reflecting greater variability in access to digital allowances across the sample.

Table 1 Descriptive Statistics for Study Variables

Variable	Mean	Standard Deviation
Digital Allowance Access	2.00	1.06
Impulsive Buying Behavior	3.22	0.83
Buying Autonomy	4.40	0.52

Pearson's correlation analysis was conducted to examine the relationships among digital allowance access, impulsive buying behavior, and buying autonomy. As shown in Table 2, digital allowance access was weakly positively correlated with buying autonomy ($r = .13$) and impulsive buying behavior ($r = .16$). Buying autonomy was negatively correlated with impulsive buying behavior ($r = -.31$). However, none of these correlations reached statistical significance, possibly due to the small sample size.

Table 2 Correlations Among Study Variables

Variable	1	2	3
1. Digital Allowance Access	—	.13	.16
2. Buying Autonomy		—	-.31
3. Impulsive Buying Behavior			—

4. Discussion

The present study examined the relationships between digital allowance access, buying autonomy, and impulsive buying behavior in adolescents. As an exploratory investigation, it aimed to identify patterns and establish relationships between them. With adolescents increasingly being exposed to digital financial tools, understanding how these tools influence buying and spending behavior and decision-making is essential. Results indicated that buying autonomy had the highest mean ($M = 4.40$), suggesting adolescents reported high confidence in making independent purchasing decisions. This likely reflects the increasing independence of adolescence. Furthermore, frequent exposure to diverse consumer choices, especially through online shopping environments, may strengthen adolescents' sense of freedom.

Impulsive buying showed a moderate mean ($M = 3.22$), indicating that adolescents engage in impulsive purchasing behaviors to some extent. This finding aligns with research suggesting that adolescents are still developing self-regulation skills, which can contribute to occasional impulsive decision-making. (Rook and Fisher; Nyrhinen et al.). In contrast, digital allowance access had a relatively low mean ($M = 2.00$), suggesting limited access among participants. This may be due to factors such as parental control over finances, cultural differences in financial independence, or age-related restrictions on digital payment tools. (Niu)

A weak positive correlation was found between digital allowance access and buying autonomy ($r = .13$). This suggests that adolescents with slightly greater access to digital allowance may feel somewhat more independent in their purchasing decisions. However, the relationship was very weak, indicating that autonomy is likely influenced more by factors such as parental trust and developmental maturity than by digital payment access alone. Similarly, digital allowance access and impulsive buying were weakly positively correlated ($r = .16$). This may suggest that easier access to digital spending tools could slightly increase impulsive buying tendencies. However, the relationship was not statistically significant, and parental monitoring may still limit impulsive spending even when digital tools are available. The most notable finding was a moderate negative correlation between buying autonomy and impulsive buying ($r = -.31$). This suggests that adolescents who perceive themselves as more independent in their purchasing decisions may actually engage in less impulsive buying. This

could be because independence promotes more thoughtful and intended decision-making. Adolescents who feel confident in their choices may be better able to regulate their spending behavior. This finding aligns with existing research indicating that autonomy can support self-regulation and responsible decision-making. (Li et al.)

It is important to note that none of the correlations reached statistical significance. This is likely due to the small sample size ($N = 20$), which reduces statistical power. As a result, these findings should be interpreted as preliminary trends rather than conclusive evidence.

These findings highlight the importance of understanding how adolescents interact with digital money. Insights from this study could inform financial literacy programs aimed at promoting responsible spending habits. Additionally, parents may play a key role in shaping adolescents' financial behaviors, particularly through monitoring and guidance.

This study has several limitations. Future research should use larger sample sizes to improve statistical power and reliability. It would also be valuable to examine differences across age groups within adolescence. Further studies could explore the role of parental monitoring and financial socialization, as well as investigate digital payment usage in more detail (Niu, 2016).

As digital financial tools become increasingly integrated into adolescents' daily lives, understanding their relationship with decision-making autonomy and impulsive buying behavior will become an increasingly important area of psychological research.

Works Cited

- Bellini, S., & Aiolfi, S. (2020). Impulse buying behavior: The mobile revolution. *International Journal of Retail & Distribution Management*, 48(1), 1–17.
<https://doi.org/10.1108/ijrdm-12-2018-0280>
- Beri, J., & Kumari, P. (2022). Digital transformation—Effect of social media on digital buying behaviour of teenagers. *Adhyayan: A Journal of Management Sciences*, 12(01), 35–40.
<https://doi.org/10.21567/adhyayan.v12i1.5>
- Bhakat, R. S. (2025). Impulse buying in the digital age: A conceptual review and research agenda toward an integrated theory of spontaneous consumption.
<http://dx.doi.org/10.2139/ssrn.5667071>
- Bhardwaj, S., Chopra, R., & Aw, E. C. X. (2026). From whims to wallets: Navigating the maze of how technology impacts impulsive buying behavior—A systematic literature review. *Internet Research*, 1–24. <https://doi.org/10.1108/intr-09-2024-1430>
- Li, Y., Chen, H., Zheng, Y. L., Wu, L. L., & Fan, C. Y. (2022). Development and validation of an autonomy questionnaire for Chinese adolescents from the perspective of network culture. *Frontiers in Psychology*, 13, 810140.
- Niu, H. J. (2017). Cyber purchasing behavior of adolescents: Family communication relationships and parental influence. *Electronic Commerce Research*, 17(3), 495–519.
<https://doi.org/10.1007/s10660-016-9223-x>
- Nyrhinen, J., Sirola, A., Koskelainen, T., Munnukka, J., & Wilska, T. A. (2024). Online antecedents for young consumers' impulse buying behavior. *Computers in Human Behavior*, 153, 108129. <https://doi.org/10.1016/j.chb.2023.108129>
- Rocha, S., Fernández, X. M., Castro, Y. R., Ferreira, S., Teixeira, L., Campos, C., & Rocha, N. B. (2023). Exploring the associations between early maladaptive schemas and impulsive and compulsive buying tendencies. *Frontiers in Psychiatry*, 14, 1157710.
- Rook, D. W., & Fisher, R. J. (1995). Normative influences on impulsive buying behavior. *Journal of Consumer Research*, 22(3), 305–313.

Digital Evidence: Technology and Oversight in Modern Courts By Joshua Kang

Abstract

Given the ever-increasing significance of digital evidence in the courtroom, the judiciary is constantly faced with the need for accuracy, reliability, and fairness of said evidence. This paper examines the collaboration between advanced technology and human experts in validating digital evidence authentication and admissibility. Based on extensive examination of contemporary literature and forensic standards, this work explores how AI, blockchain preservation, and cryptographic hashing contribute to secure evidence collection, and analyzes emerging issues of deepfakes, encryption, and cross-jurisdictional cloud storage. This work argues that without verified software and tools, established protocols, and unbiased human discretion, unchecked and rapid technological development is inherently likely to lead to wrongful and/or unfair convictions. This paper advocates for legislation in the form of compulsory validation of forensic software, modernizing MLATs, and requiring authentication standards for deepfakes. These recommendations are specifically targeted at a policy-making panel of judges, will be conveyed in a formal presentation, and used as a basis for expert discussion and evaluation.

Keywords

Digital Evidence, Digital Forensics, Blockchain, Cryptographic Hashing, Artificial Intelligence, Deepfakes, Cloud Forensics, MLAT, Courtroom Authentication, Forensic Tool Validation.

Introduction

Digital technology has dramatically impacted litigation today. Digital data now operates as a core element of both civil and criminal proceedings. Digital phenomena such as emails, text messages, GPS logs, surveillance footage, and cloud-based documents are now used as evidence of probative truth. With such widespread utility of digital evidence, courts are demanding enhanced evaluation and validation for such evidence before acceptance as a valid body of evidence. Forensic tools must be tested for validity and reliability. Otherwise, digital evidence can be disregarded and denied entry into the court as invalid (Brunty). Courts require such technologies to not only be validated but also scientifically accurate to promote procedural fairness, and it must further run under the supervision of an expert to establish credibility and reliability (Grimm et al.).

The amount of digital evidence is increasing and has become a staple of modern court proceedings. By one estimate, digital evidence is a factor in approximately 90% of criminal cases in the United States (Miller). The admissibility and effectiveness of digital evidence depend on the skillful integration of advanced forensic technologies and expert human analysis. Without validated tools, established methodologies, and transparent interpretation, evolving technologies such as AI, cloud storage, encryption, and blockchain risk producing compromised and unjust

courtroom outcomes. Table 1 summarizes the primary types of digital evidence now appearing in court, their common uses, preservation methods, and associated challenges.

Table 1. Types of digital evidence in modern litigation: uses, preservation methods, and key challenges.

Type of Digital Evidence	Common Use in Court	Preservation Method	Key Challenge
Emails & Text Messages	Intent, communication timelines	Cryptographic hashing	Metadata authenticity
GPS & Location Data	Suspect placement, alibi verification	Device forensic imaging	Third-party data access
Surveillance Footage	Identity, timeline of events	Chain of custody logs	Deepfake manipulation
Cloud-Based Documents	Financial records, communications	Blockchain ledger	Cross-jurisdictional access
Social Media Posts	Statements, relationships, motive	Screenshot + hash verification	Authenticity, account hacking
Encrypted Messages	Criminal coordination, conspiracy	Metadata & traffic analysis	End-to-end encryption barriers

Note. Compiled from Brunty, Karagiannis and Vergidis, Sarhan et al., and Patil et al.

Technological Tools for Digital Evidence Authentication

The developments in technology have improved digital forensic capabilities considerably. AI has the power to process enormous datasets and discern patterns more quickly than a human investigator (Hildebrandt and Kunz). The development of cloud computing offered convenient storing and retrieving of evidence, but created problems of multi-tenancy, volatility of data, and jurisdiction (Karagiannis and Vergidis). The use of blockchain-based preservation systems and cryptographic hashing provides a time-resistant chain of evidence that would safeguard the evidence (Patil et al.). Even with these advances, courts continually require new standards of law and expert explanation of results for their proper understanding (Harris).

Blockchain-Based Preservation

The courts today are finding technology to be crucial to maintaining control over the process through which digital evidence is collected, preserved, and authenticated. The nature of digital evidence is such that as its use becomes more widespread, the courts must find ways of ensuring its integrity and indelibility. A blockchain system can record every bit of evidence from collection all the way to introduction as evidence before the courts in an indelible and secure manner. Since the system consists of multiple identical digital copies of each ledger, it is impossible to tamper with a particular piece of evidence without breaking the integrity of the entire system (Patil et al.). By maintaining a transparent and verifiable chain of custody, it will be clear whether evidence has been altered, since even the smallest of alterations could impact the entire case.

Numerous real-world scenarios demonstrate the usefulness of blockchain in court. For instances of financial fraud, blockchain has been utilized in tracing transactions and making indisputable timelines of all events, allowing the court to examine timestamped digital evidence instead of relying on the oral testimony of witnesses and their sometimes-faulty memories (Patil et al.).

Cryptographic Hashing

Cryptographic hashing is another important feature that can ensure the authenticity of digital evidence. This is the process by which a digital fingerprint of each file is generated, and if any change is made to that file, the digital fingerprint will be altered as well. The use of this technique ensures the authenticity of digital evidence by comparing initial and current hash values, providing a binary and non-subjective means of verifying any manipulation of evidence (Liu and Zheng). In addition, the use of this process along with a blockchain system creates a highly secure environment for data storage and authentication.

Artificial Intelligence in Forensic Review

Artificial intelligence enhances the courts' ability to search and verify digital evidence. Manually searching data would be an enormous task for any forensic investigator, as it involves an overabundance of digital information including emails, financial transaction reports, and communication data. Artificial intelligence can easily and rapidly go through large sets of information and search for unusual patterns to locate tampering. It is exceptionally useful in finding discrepancies within multiple sources of data and can identify subtle inconsistencies that an investigator may not notice (Hildebrandt and Kunz). Additionally, AI verifies the metadata of a document, including information about how and when evidence was downloaded, copied, or modified.

Many cases of wrongful convictions have been reversed once the modification of a digital file was revealed using artificial intelligence in forensics (Hildebrandt and Kunz). Blockchain systems, cryptography, and artificial intelligence provide a valuable service in the legal community for ensuring the safety and verification of digital evidence, but it is imperative to use them within established legal guidelines and with a qualified professional overseeing their use.

The Necessity of Forensic Tool Validation

Because digital evidence is tied to forensic techniques, test validation becomes an important part of proving that evidence is reliable and fair. Digital evidence collected with tools that have not been tested and validated will often be rejected by the court (Brunty). Without testing and documented validation, the judge cannot be certain that a process or tool will achieve desired results, or that results accurately reflect the evidence.

Validation is an especially important concept in AI-driven digital forensics. Artificial intelligence algorithms are highly complex (Hildebrandt and Kunz), and a significant difference

between the real world and a model based on training of incomplete or poorly reviewed corpora can cause normal artifacts to look suspicious, or vice versa. Validation explains the way AI is used by providing reliability error rates in varied conditions, along with detailed explanations of limitations. Judges want to confirm that they are not accepting AI findings as facts simply on the basis of algorithmic output alone.

Validation brings credibility because it allows judges and juries to examine human input. The legal system relies on both technology and human expertise (Grimm et al.). Validated software provides findings that must be reviewed by a qualified analyst who can detect inconsistencies, convert technical terminology into comprehensible terms, and determine the legal significance of findings (Miller). Ignoring tool validation will also harm judicial proceedings. In court, unvalidated evidence is subject to adversarial challenges, and courts may throw out or delay proceedings as a result. Unvalidated software can produce timelines that could imply motive and presence where they do not exist, placing the presumption of innocence in jeopardy (Brunty).

Emerging Challenges: Deepfakes, Encryption, and Cloud Jurisdiction

Deepfakes and the Deepfake Detection Crisis

When defendants contest the authenticity of audio or video recordings, courts must decide who carries the burden of proving authenticity. No clear unified standard has emerged, and many courts have approached this issue without consistent standards of proof, leaving the matter to individual jurisdictions (Ulbricht et al.). Deepfakes make it significantly harder for courts to verify digital evidence (Runyon). Unlike other methods of alteration, they may leave behind no trace and be difficult for an expert to recognize. This could have dire implications in court, as falsified evidence may influence a jury or otherwise be assumed genuine.

The challenge is compounded by the significant performance gap between laboratory testing and real-world deployment of detection tools. As shown in Table 2, automated deepfake detection systems experience an accuracy drop of 45-50% when confronted with modern, real-world deepfakes compared to controlled lab benchmarks (Chandra et al.). Most alarmingly, unaided human detection of high-quality deepfake video is only approximately 24.5% accurate (Diel et al.), meaning that without forensic expert assistance, juries are highly likely to be deceived by fabricated media. Between 2023 and 2024, there was a fourfold increase in deepfakes detected in fraud cases, and deepfake identity fraud attempts surged 3,000% in 2023 alone (Chandra et al.).

Table 2. Deepfake detection performance: laboratory vs. real-world accuracy (2024-2025).

Detection Method	Lab AUC Score	Real-World AUC	Performance Drop	Source
Open-Source AI Models (avg.)	~90%	~45-50%	~45-50%	Chandra et al., 2025

Commercial Detection Systems	~85%	~65-70%	~15-20%	Chandra et al., 2025
Human Forensic Analysts	~90%	~90%	~0%	Chandra et al., 2025
Unaided Human Detection	N/A	~24.5%	N/A	Diel et al., 2024

Note. AUC = Area Under the Receiver Operating Curve. Data from Chandra et al. and Diel et al.

Encrypted Communication

Another large hurdle to the authentication of digital evidence comes from encrypted communication. A large number of communication platforms currently utilize end-to-end encryption, prohibiting investigators' access to the content of messages. Instead, forensic experts rely on metadata and traffic analysis to determine what actions took place and who participated. Metadata is not always indicative of all details (Sarhan et al.). Thus, it does not always tell the whole story, and it can be difficult to assert what actions an individual took and their intention in doing so. Furthermore, widespread use of encryption raises ethical dilemmas that have consequences for its use as legal evidence.

Cloud Storage and Cross-Jurisdictional Challenges

The widespread use of cloud systems raises an additional set of challenges related to jurisdiction and evidence acquisition. Digital evidence is often stored across multiple cloud servers residing in various countries, each with its own legal jurisdiction and data protection laws. Acquiring evidence located in the cloud can be a taxing process as it may require navigating international legal statutes (Karagiannis and Vergidis). In a multi-tenant cloud system, in which several people access the same server, the question arises as to whether one's data is completely separated from that of other tenants, casting doubt on whether evidence will have been securely protected from corruption.

Mutual Legal Assistance Treaties (MLATs), by which governments cooperate with other states on criminal investigations, predate cloud computing and are often described as too slow and cumbersome for contemporary digital investigations. A formal MLAT request can take months or even years to process, and evidence contained in the cloud is often volatile and can be destroyed or moved to another jurisdiction before a request is fulfilled (Karagiannis and Vergidis). This leaves a gaping hole between the speed of cybercrime and the speed of the legal system.

Discussion

The ongoing complexity of digital threats clearly calls for a combination of technological advancement and legal regulation in order to protect due process. Digital evidence plays a key role in influencing judicial outcomes, but its evolving complexity has brought new challenges and risks to authenticity, reliability, and fairness. Courts consistently rely on human experts to

validate forensic tools, interpret technical findings, and explain methodologies within established legal standards. However, without consistent validation, transparent procedures, and human oversight, digital evidence will continue to be vulnerable to corruption, misinterpretation, and unfair influence on verdicts. Table 3 summarizes the primary challenges identified in this review and proposes corresponding reforms.

Table 3. Summary of digital evidence challenges and proposed legislative reforms.

Challenge Area	Current Status	Risk to Justice	Proposed Reform
Forensic Tool Validation	No mandatory federal standard	Wrongful convictions from unvalidated software	Legislate compulsory tool validation
Deepfake Detection	Automated tools drop ~45% in real-world accuracy	Juries may convict on fabricated media	Require third-party expert authentication
Encrypted Communications	Metadata used as substitute; incomplete picture	Misattribution of intent or action	Develop legal metadata standards
Cross-Border Cloud Evidence	MLATs take months/years; predate cloud era	Evidence loss through volatility or destruction	Modernize MLATs for digital evidence

Note. Compiled from Brunty, Ulbricht et al., Karagiannis and Vergidis, and Sarhan et al.

Judges and legislators must quickly introduce new methods for dealing with the expanding issue of digital evidence. Standards of validation must be legislated so that any forensic software brought into the courtroom can be examined and proven to be thoroughly tested and valid. Similarly, mutual legal assistance treaties must be reformed to adequately reflect the transient, cross-border nature of the cloud. When considering deepfakes, judges should require third-party independent verification of all audio and video evidence submitted, so as to provide minimum across-the-board standards. If new procedures are not put in place, technology and liability will continue to outpace one another, and because technology increasingly forms the backbone of courtroom decision-making, wrongful convictions, case failures, and compromised due process will occur with greater frequency.

The system depends on the courts to keep pace with the technology on which they rely. New validation procedures and requirements, uniform testing, and the role of expert evaluation are not just optional improvements. They are imperative for justice. As digital evidence continues to grow in prevalence and complexity, the legal system must establish standardized frameworks that integrate both technological tools and human expertise to ensure that every piece of evidence admitted into the courtroom is authentic, reliable, and fairly evaluated.

Works Cited

- Brunty, Josh. "Validation of Forensic Tools and Methods: A Primer for the Digital Forensics Examiner." *WIREs Forensic Science*, vol. 5, no. 2, 2023, article e1474, <https://doi.org/10.1002/wfs2.1474>.
- Chandra, Naveen A., et al. "Deepfake-Eval-2024: A Multi-Modal In-the-Wild Benchmark of Deepfakes Circulated in 2024." *arXiv*, 2025, <https://doi.org/10.48550/arXiv.2503.02857>.
- Diel, Alexander, et al. "Human Performance in Detecting Deepfakes: A Systematic Review and Meta-Analysis of 56 Papers." *Computers in Human Behavior Reports*, vol. 16, 2024, article 100538, <https://doi.org/10.1016/j.chbr.2024.100538>.
- Grimm, Paul W., et al. "Authentication of Social Media Evidence." *American Journal of Trial Advocacy*, vol. 36, 2013, pp. 433–72, scholarship.law.duke.edu/faculty_scholarship/4418.
- Harris, Mark. "Blockchain-Based Evidence in Courts: Standards, Reliability, and Admissibility Challenges." *Legal Studies in Digital Age*, vol. 2, no. 3, 2023, pp. 49–63, jlsda.com/index.php/ljsda/article/view/321.
- Hildebrandt, Mario, and Christopher Kunz. "Digital Forensics AI: Evaluating, Standardizing and Optimizing Digital Evidence Mining Techniques." *KI - Künstliche Intelligenz*, vol. 36, no. 3, 2022, pp. 281–91, <https://doi.org/10.1007/s13218-022-00763-9>.
- Karagiannis, Vasileios, and Konstantinos Vergidis. "Digital Evidence and Cloud Forensics: Contemporary Legal Challenges and the Power of Disposal." *Information*, vol. 12, no. 5, 2021, p. 181, <https://doi.org/10.3390/info12050181>.
- Liu, Shuli, and Qun Zheng. "A Study of a Blockchain-Based Judicial Evidence Preservation Scheme." *Blockchain Research and Applications*, vol. 5, no. 2, 2024, article 100192, <https://doi.org/10.1016/j.bcra.2024.100192>.
- Miller, Craig M. "A Survey of Prosecutors and Investigators Using Digital Evidence: A Starting Point." *Forensic Science International: Synergy*, vol. 6, 2023, article 100296, <https://doi.org/10.1016/j.fsisyn.2022.100296>.
- Patil, Harshwardhan, et al. "Potential Applicability of Blockchain Technology in the Maintenance of Chain of Custody in Forensic Casework." *Egyptian Journal of Forensic Sciences*, vol. 14, no. 1, 2024, <https://doi.org/10.1186/s41935-023-00383-w>.
- Runyon, Nadya. "Deepfakes on Trial: How Judges Are Navigating AI Evidence Authentication." Thomson Reuters Institute, 8 May 2025, www.thomsonreuters.com/en-us/posts/ai-in-courts/deepfakes-evidence-authentication/.
- Sarhan, Sara A. E., et al. "A Framework for Digital Forensics of Encrypted Real-Time Network Traffic, Instant Messaging, and VoIP Application Case Study." *Ain Shams Engineering Journal*, vol. 14, no. 9, 2023, article 102069, <https://doi.org/10.1016/j.asej.2022.102069>.
- Ulbricht, Beth R., et al. "Digital Eyewitnesses: Using New Technologies to Authenticate Evidence in Human Rights Litigation." *Stanford Law Review*, vol. 74, 2022, pp. 851–935, review.law.stanford.edu/wp-content/uploads/sites/3/2022/04/Ulbricht-et-al.-74-Stan.-L.-R-ev.-851.pdf.

Autonomous Driving Systems: Principles, Technologies, and Functioning By Taleyn Khan

Abstract

The rapid development of artificial intelligence (AI) and automation technologies has positioned autonomous vehicles as a transformative innovation in modern transportation systems. This study aims to understand the operational mechanisms of autonomous vehicles in real-world contexts and to examine the limitations associated with their deployment. Data were collected through structured interviews with engineers working in autonomous vehicle technology. An assertion–reason–evidence analytical approach was employed to interpret participant responses. Findings indicate that autonomous vehicles rely heavily on sensor fusion, AI, and large-scale data processing for navigation and decision-making. However, environmental conditions, unpredictability in human behavior, and ethical decision-making challenges remain significant barriers. The study concludes that while autonomous vehicles offer improved safety and efficiency, they are not yet fully reliable across all contexts. Their successful integration will depend on balancing automation with human oversight and addressing key technological and ethical concerns.

Keywords Autonomous Vehicles, Artificial Intelligence, Sensor Fusion, Machine Learning, Self-Driving Systems.

Introduction

The field of transportation has experienced revolutionary changes over the past century, ranging from simple horse-drawn means to advanced electric technology. Despite these developments, road accidents have remained a major problem in the world, with the major cause of accidents being attributed to issues related to human error. This problem was addressed by engineers and scientists in the design of autonomous cars that do not rely on a driver. Autonomous cars use sensors, artificial intelligence, and real-time data analysis to operate. Firms such as Tesla, Waymo, and Google have invested a lot in developing cars that have the capability to sense their surroundings and take appropriate driving directions on their own. The cars are made up of components such as cameras and artificial intelligence systems to facilitate smooth movements. (Fagnant & Kockelman, 2015)

Interest in autonomous vehicles has been growing consistently in recent years due to concerns with road safety and traffic management. The number of road accidents runs into millions every year, with a large number of them caused by faulty human decisions based on distraction, fatigue, or poor judgment. Self-driving technology aims at restricting these problems by allowing the vehicle to make accurate and continuous decisions through computer systems. Most of the major technology and automobile companies, such as Tesla, Waymo, and Google, are therefore involved in actively developing systems that can allow the car to sense the environment, follow the traffic rules, and drive safely without any human intervention. The rapid

growth in this regard is hence an indication of the increasing importance of autonomous transportation.

Previous studies have shown that autonomous vehicles can significantly reduce accident rates by minimizing human mistakes. Researchers have found that sensor-based systems such as cameras, radar, and LiDAR allow vehicles to detect obstacles and react faster than human drivers. (Fagnant & Kockelman, 2015). Other research highlights that artificial intelligence and machine learning help cars continuously improve their driving performance through data analysis. However, findings also indicate concerns related to safety reliability, ethical decision-making, and system security.

Literature review

Existing literature highlights that autonomous vehicles are complex systems integrating sensing technologies, artificial intelligence, and control mechanisms. Most literature has reported only individual technologies, such as sensors and artificial intelligence, but not so much about the whole working process of autonomous vehicles.

Moreover, there is a scarcity of research that offers simplified and holistic descriptions of the ideas, enabling readers like students to understand the step-by-step operating process of these machines. The available literature is highly technical and not easy to understand for anyone other than the experts. Moreover, there is a need to conduct further research that effectively describes the basic operating mechanism of autonomous cars and divides the complex operating processes into easily understandable steps.

Filling this information gap will not only enhance the general public's awareness of autonomous functionalities, but will also promote an appreciation for what this technology can and cannot do for them.

The purpose of this study is to seek an understanding of how autonomous cars are able to function by evaluating the fundamental technologies employed to facilitate successful self-driving through various components such as sensors, artificial intelligence, data, data processing, decision-making systems, and vehicle control systems. This specific aim of this study is to provide an overall framework of how an autonomous vehicle is able to function while at the same time showing the overall interaction between different components to provide safe, acceptable, and intelligent driving systems for the future, thus enhancing our overall understanding of advanced technologies about the future of smart cars.

Through the research papers reviewed on autonomous vehicles (self-driving cars), the authors generally define autonomous driving as a complex and interdisciplinary system that combines sensing technologies, artificial intelligence, control engineering, and safe-critical decision-making. The papers generally highlight that autonomous cars do not depend on one technology but on a layered system that combines perception, decision-making, and vehicle control in real-time.

One of the key findings from the literature is that autonomous vehicles seek to minimize human error, which is accountable for the majority of road accidents. The authors generally point

out that machines can focus on a task continuously, respond faster than humans, and adhere to traffic rules. However, the papers also generally point out that full autonomy is a complex challenge because of unpredictable environments, human behavior, and ethics. In general, the research literature portrays autonomous cars as a developing technology that has made substantial progress in controlled settings but still needs extensive validation, regulation, and public acceptance.

One of the major themes that run through the research papers is that autonomous vehicles depend on sophisticated sensing technology in order to continuously perceive and interpret their environment. The researchers make it clear that autonomous vehicles do not depend on vision, as human drivers do, but instead use a variety of sensors such as cameras, LiDAR, radar, GPS, and inertial sensors to gather data about the environment, including road conditions, other vehicles, pedestrians, traffic signals, and obstacles. Each of these sensors provides a different kind of data, such as visual information, distance information, or relative speed information. The theme of the research papers is that accurate perception of the environment is the key to autonomy, since all other decisions are dependent on the quality of the sensed data.

Another important theme that is widely discussed in the literature is sensor fusion. Sensor fusion is the process of combining data from multiple sensors to form a single perception of the environment. The literature suggests that it is not possible for a single sensor to offer complete and accurate information about the environment at all times. For instance, cameras provide detailed visual information but are not accurate in terms of depth perception, while LiDAR sensors are accurate in terms of distance measurement but do not offer semantic data. The literature suggests that sensor fusion helps to combine the complementary advantages of different sensors to improve accuracy and consistency. The theme emphasizes that autonomous vehicles require not only advanced sensors but also sophisticated algorithms for data integration. The literature suggests that sensor fusion is important for safe navigation in complex traffic environments where rapid changes occur. The literature suggests that sensor fusion is an important mechanism that helps autonomous vehicles to maintain accurate perception and decision-making capabilities even in uncertain and challenging environments.

The use of artificial intelligence and machine learning is one of the themes that appear in the research papers. The authors explain that learning algorithms, especially deep learning algorithms, are employed to interpret sensor data and detect patterns in the driving environment. These algorithms allow autonomous vehicles to detect objects like pedestrians, cars, traffic lights, and road signs with high accuracy. In addition to object detection, AI algorithms are also employed to classify behaviors and predict future actions of other road users. The papers highlight that machine learning enables autonomous vehicles to operate in complex and dynamic environments that cannot be completely represented by rule-based programming. However, the authors also admit that there are challenges in using AI, such as bias, lack of transparency, and robustness to unexpected events. This theme indicates that although AI greatly improves the ability of autonomous vehicles, it also raises new issues about reliability and accountability. In

general, the papers indicate that artificial intelligence is a great enabler of autonomous vehicles but needs to be properly trained and supervised.

One of the themes that have been consistently presented in the literature is the complexity of autonomous decision-making and trajectory planning. According to the literature, autonomous vehicles are required to make decisions on how to behave in a dynamic traffic environment while following traffic rules and ensuring safety. Autonomous decision-making involves making decisions on the most suitable driving actions, such as turning, braking, or yielding, while trajectory planning involves computing smooth and collision-free paths. The theme is emphasized in the literature by pointing out that these decisions have to consider uncertainties such as human behavior and road conditions. Unlike human drivers, autonomous vehicles have to model these uncertainties using mathematical models and optimization algorithms. This theme underlines the fact that real-time decision-making is one of the most difficult tasks in autonomous vehicles because of the strict time constraints and safety considerations. According to the literature, inappropriate decision-making can result in unsafe or uncomfortable driving actions.

The research papers always emphasize the importance of control systems in carrying out the autonomous driving decisions. After the trajectory has been planned, the control algorithms are tasked with the responsibility of translating the plan into actual steering, acceleration, and braking maneuvers. The researchers always point out that these systems must be able to guarantee the stability of the vehicle, comfort of the passengers, and reaction to external disturbances such as road friction or sudden appearance of obstacles. This topic proves that autonomy is not just about high-level intelligence but also about low-level execution. Even the slightest mistake in control systems can jeopardize safety, irrespective of how sophisticated the perception or planning systems are. The literature also covers feedback systems that make continuous adjustments to control actions based on real-time sensor data. This topic always emphasizes that control systems are the final and crucial link between autonomous decision-making and actual vehicle motion.

The set of research papers reviewed together portray autonomous cars as highly complex and integrated systems that combine sensing technologies, artificial intelligence, decision-making algorithms, and control systems to accomplish driving functions with minimal human intervention. The thematic analysis of the papers clearly indicates that autonomous driving does not rely on a single innovation but rather on the collective and coordinated operation of multiple subsystems. Environmental awareness begins with accurate perception through advanced sensors and reliable sensor fusion. Artificial intelligence and machine learning further improve this capability by allowing interpretation, prediction, and adaptive responses in dynamic traffic environments.

The literature also highlights the importance of decision-making, trajectory planning, and control systems for ensuring safe and optimal vehicle operation. Nevertheless, all researchers agree that mere technological progress is not enough for the successful implementation of autonomous cars. Rather, safety assurance, ethical responsibility, legal accountability, and social

acceptance must be simultaneously addressed and managed. In summary, the papers reviewed suggest that autonomous cars have immense potential for minimizing traffic accidents and optimizing transportation efficiency, but their successful implementation requires rigorous validation, proper design, and alignment with ethical and legal frameworks. In other words, autonomous driving is best understood as an evolving socio-technical system rather than a purely automated replacement for human drivers.

METHODOLOGY

This study adopts a qualitative research design to explore the functioning and limitations of autonomous vehicles through expert perspectives. A purposive sample of five engineers working in the field of autonomous vehicle technology was selected. Participants were chosen based on their expertise and professional experience in AI-driven transportation systems. Data was collected through structured interviews. Interviews were conducted virtually and responses were documented for analysis. An assertion–reason–evidence (ARE) framework combined with thematic analysis was used.

Result table

Participant Quote	Code	Theme
“detects environment through cameras, radar, LiDAR”	Sensor dependency	Perception & Sensor Systems
“sensor fusion improves accuracy”	Sensor fusion	Perception & Sensor Systems
“performance compromised in fog or rain”	Environmental limitation	Environmental limitation
“human behavior is unpredictable”	Human unpredictability	Human Behavior Challenges
“large data sets are required”	Data dependency	Data & AI Dependence
“labeling data is challenging”	Data labeling challenge	Data & AI Dependence
“minor errors lead to wrong decisions”	System error	Decision-Making & Safety
“system tries to minimize harm”	Safety prioritization	Decision-Making & Safety
“gradual adoption expected”	Gradual adoption	Development Challenges

Discussion

This research was designed to identify how autonomous cars work and their practical difficulties from the perspective of experts in the field. Data was obtained from interviews and thematic analysis.

1. Perception and Sensor Systems

Another major theme that was picked up from the conversation was the significance of perception systems in autonomous vehicles. According to the engineers, self-driving cars depend significantly on sensing systems like cameras, radars, and LiDAR to sense their environment. In fact, there are systems in place to combine the input from various sensors through what is termed sensor fusion.

However, how successful these systems are can be gauged by the extent of integration between different sensor inputs. “Sensor fusion helps improve accuracy since no one sensor is perfect,” explained a participant.

2. Environmental and Real-World Limitations

Environmental effects on the system were another important point raised by the engineers. They noted that the environment brings about variation that cannot be controlled. Fog, rain, and poor visibility can have a great effect on the sensors.

It was noted that “performance may still be compromised in heavy rain or fog.” From this statement, it can be seen that despite all the technological advancements made, the environment remains one of the greatest obstacles to achieving full autonomy.

3. Human Behavior Challenges

The unpredictable nature of human beings became a significant issue. Unlike the laboratory, where conditions can be controlled, driving in the real world includes interaction with other drivers and pedestrians who are unpredictable in their behavior.

As pointed out by the engineer, “Human behavior is unpredictable.” This issue limits the ability of autonomous systems to make decisions based on anticipated behavior.

4. Data and AI Dependence

Autonomous driving is highly dependent on artificial intelligence algorithms trained using massive amounts of data. An important theme that emerged from the research process was that of the dependence on good-quality data for effective performance.

As was said earlier, “the main difficulty is the collection of large amounts of data and their labeling.” This indicates the direct correlation between the efficiency of an AI system and the data used to train it.

5. Decision-Making and Safety

Safety is an important feature of any self-driving car system. The programming is made to be focused on making decisions that result in the least damage to the driver and passengers. But, as can be observed, the issues of ethics still persist and are yet to be solved.

According to the engineer, “The system will try to do as little harm as possible,” but then added, “It’s complex to program ethics into machines.”

6. Development and Adoption Challenges

Lastly, problems associated with the widespread deployment of autonomous vehicles were also pointed out in the study. The engineers noted that adoption would be gradual because of several reasons, both technological and environmental in nature.

They mentioned that “gradual adoption should be expected in the coming years.”

Conclusion

The objective of this study was to determine how autonomous cars work and to analyze their performance in the real world by drawing information from experts in engineering. The results show that the operation of autonomous driving technology involves a series of processes such as perception, computation, forecasting, and control, which mainly depend on artificial intelligence and sensing technologies. Autonomous driving technology uses sensor fusion extensively, which involves the combination of visual input from cameras, radar, and light detection and ranging (LiDAR).

The study also reveals that there are still some constraints that exist, even though there have been remarkable improvements. Some environmental issues, including fog, rain, and poor lighting, may lead to inaccurate data in the sensors and thus decrease the reliability of the systems. The most important constraint that has been mentioned in this study is that the behavior of humans cannot be predicted. Contrary to a controlled environment, driving in real-life situations will involve the interaction of human beings, which cannot be predicted easily. Another crucial factor that comes to light during this study is the reliance on vast amounts of data for training the artificial intelligence systems. The problem of gathering data and ensuring it is of good quality and properly labeled still stands out as one of the biggest problems. In addition, the existing technologies operate as sophisticated pattern recognition devices that work well in a known environment, yet may face problems when dealing with an unknown environment. Making ethical decisions is another problem that cannot be easily solved. Considering the implementation, from the research findings, the use of self-driving cars is likely to be a stepwise process rather than immediate. This may take into account issues like technological challenges, infrastructure needs, legal concerns, among others. Although self-driving cars can help to minimize traffic accidents as well as boost efficiency in transport systems, they still cannot completely take over from human drivers.

From all the above, it can be seen that self-driving cars constitute a remarkable technological leap, which has the capacity to revolutionize the way current transportation systems function. Nonetheless, in order for them to be effective and beneficial, it is necessary for constant innovations in technology and adaptation to external factors, as well as incorporating human and ethical issues, to take place.

Works Cited

1. Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. Transportation Research Part A. <https://www.sciencedirect.com/science/article/pii/S0965856415000804>
2. Research Blog. (2020). Self-driving car technology and AI systems. <https://ai.googleblog.com/search/label/Self-Driving%20Cars>
3. Parekh, D., Poddar, N., Rajpurkar, A., Chahal, M., Kumar, N., Joshi, G. P., & Cho, W. (2022). Autonomous driving systems: Overview and challenges. <https://www.mdpi.com/2079-9292/11/14/2162>
4. W. (n.d.). Autonomous vehicle intelligence systems. <https://injm.com/index.php/ijifi/article/view/209>
5. Fayyad, J., Jaradat, M. A., Gruyer, D., & Najjaran, H. (2020). Deep learning sensor fusion for autonomous vehicles. <https://pmc.nih.gov/articles/PMC7436174/>
6. Teng, S., Hu, X., Deng, P., et al. (2023). Autonomous driving systems review. <https://arxiv.org/abs/2303.09824>
7. Paden, B., Čáp, M., Yong, S. Z., Yershov, D., & Frazzoli, E. (2016). A survey of motion planning. <https://arxiv.org/abs/1604.07446>

The Influence of Cognitive Behavioral Therapy and Mindfulness-Based Interventions on Executive Functioning and Self-Regulation for Individuals with Anxiety By Ishaan Reddy

Abstract

Worldwide, anxiety symptoms and disorders affect a large proportion of individuals through several impairments of executive functioning or cognitive processes that facilitate goal-oriented behavior. These include thought formation, selective attention, cognitive flexibility, and task performance. Additionally, anxiety can hinder the ability to manage behaviors, emotions, and impulsivity. The purpose of the literature review is to analyze the effects of cognitive behavioral therapy (CBT) and mindfulness-based interventions on executive functioning and self-regulation for individuals with anxiety. Moreover, this review examined seven peer-reviewed studies to compare and contrast the effectiveness of cognitive behavioral therapy and mindfulness-based interventions such as Mindfulness-Based Stress Reduction in providing coping mechanisms and minimizing the effects of anxiety. Across all findings, both CBT and mindfulness-based interventions not only eased the intensity of anxiety symptoms, but proved supportive of cognitive and emotional domains. Specifically, CBT reversed the formation of maladaptive thoughts, strengthening comfortable adaptation to change in the environment, and providing them with the tools for approaching anxiety-arousing events. On the other hand, mindfulness-based interventions promoted the maintenance of positive self-awareness, attention to body sensations, and the practice of relaxative methods that reduced the frequency of impulsive behaviors and emotional fluctuations. Although both treatments differed in their mechanisms, they comparably generated beneficial influence on enhancing the processes behind executive functioning, self-regulation, and reducing anxiety symptoms in a way that reinforces psychological well-being. Overall, the findings suggest that both treatments represent promising approaches that can be applied by clinicians and coincide with an individual's preferences to ensure satisfactory results.

Introduction

According to the World Health Organization, 359 million people worldwide were living with an anxiety diagnosis in 2021 and roughly 4.4% of the current global population are affected by an anxiety disorder (WHO, 2025). The commonality of anxiety disorders has slowly been rising and increasing the likelihood of developing other commonly comorbid conditions. Anxiety is a complex mental state of fear and uneasiness toward a certain stimuli, which includes symptoms such as heightened heart rate, lower concentration, irritability, and physical and mental tension (WHO, 2025). Anxiety can be diagnosed as specific disorders (e.g. generalized anxiety disorder, social anxiety, panic disorder, etc.), that are associated with several negative social outcomes, which include senses of well-being, self-esteem, and formation of enduring social connections. However, recent research has viewed heightened levels of anxiety and its related disorders as conditions that negatively impair cognitive, behavioral, and self-regulatory processes (WHO, 2025). Specifically, the cognitive aspects focus on executive functioning which

is an umbrella that encompasses higher-order cognitive abilities that enable behavioral flexibility, adaptive working memory, and formation of goal-oriented behaviors. These serve as the foundation for self-regulatory mechanisms that manage impulses and strengthen behaviors that can satisfy basic needs and desires in a way that guarantees long-term satisfaction to a certain extent. Despite these negative symptoms and disturbances to daily life, it is believed by evolutionary psychologists that anxiety played an important role in human evolution. It has been proposed that anxiety is a defense mechanism generated over time through evolution of our ancestors to keep safe from danger, but heightened persistence of anxiety and possible culmination into disorders can negatively interfere with daily functioning across several social environments (WHO, 2025).

One of the primary focuses of current anxiety research is its influence on executive functions and self-regulation (Cleveland Clinic, 2024). Executive functions are defined as cognitive processes that enable people to generate and incorporate goal-oriented behavior, as well as adjust selective attention and working memory capacities in situations requiring engagement (Cleveland Clinic, 2024). Self-regulation is linked to executive functioning in terms of balancing emotional fluctuations, behaviors, and thought processing (Cleveland Clinic, 2024). Additionally, research about anxiety's disruption of executive functioning and self-regulation provides further information about common psychological treatments of anxiety: cognitive behavioral therapy and mindfulness-based interventions. Cognitive behavioral therapy is a therapeutic approach that encompasses several smaller, yet related interventions that are utilized to restructure thought processes and reduce the occurrence of maladaptive behaviors related to anxiety (WHO, 2025). By contrast, mindfulness-based interventions focus on reconnection with the present moment, self-awareness, and relaxation of bodily functions. Prominent examples include Mindfulness-Based Stress Reduction and Mindfulness-Based Cognitive Therapy that incorporates meditation, yoga, deep breathing, and other calming techniques that can exhibit a positive influence on executive functioning and self-regulation for those with high levels of anxiety or those with a diagnosis (American Psychological Association, 2019). Understanding the means by which cognitive behavioral therapy and mindfulness are implemented can strengthen the understanding of the effectiveness of these common psychological treatments in managing anxiety symptoms.

Therefore, the purpose of this literature review is to analyze published research on efficacy of cognitive behavioral therapy and mindfulness on anxiety and answer the research question how does cognitive behavioral therapy and mindfulness-based interventions impact executive functioning and self-regulation in individuals with anxiety.

Methods

A literature review was conducted by identifying peer-reviewed studies that examined the influences of cognitive behavioral therapy and mindfulness-based interventions on executive functioning and self-regulation for individuals with anxiety or its associated disorders. The literature search was conducted in February, 2026 using two scholarly databases: Google Scholar

and PubMed. The inclusion criteria were studies that involved the use of Cognitive Behavioral therapy or mindfulness-based interventions, executive functioning or related cognitive mechanisms, self-regulation, and a population diagnosed with anxiety disorders or measured with high levels of symptoms. The exclusion criteria were studies that did not utilize one of the two interventions, did not analyze cognitive, self-regulatory, and behavioral variables, and did not include a population with certain forms of anxiety.

Results

Following 20 full text reviews, a total of 7 peer-reviewed studies were collected for analysis in the literature review. In total, the studies contained a diverse group of participants with different diagnoses. The included studies also used a variety of measurement tools of anxiety symptoms, executive function abilities, and self-regulation skills. There was also variation in study design (e.g., review, meta-analyses, randomized controlled trials, and direct comparisons). In brief, one of the included studies focused on the relationship between anxiety and self-regulation of behaviors and emotions, and executive functioning and associated cognitive abilities; two studies focused on the influence of cognitive-behavioral therapy on executive functioning; two studies specifically analyzed the effects of mindfulness-based interventions, such as Mindfulness-Based Stress Reduction programs and Mindfulness-Based Cognitive Therapy; and two studies compared the influences of the two psychological treatments.

Interplay between Anxiety and Cognition and Self-Regulation

Anxiety comprises several underlying mechanisms that influence cognitive and self-regulatory processes. A systematic review was conducted by Rabner and colleagues in 2024 to examine the resulting impairments anxiety can cast upon these cognitive elements and additional domains. The inclusion criteria were such that included studies needed to have a sample of individuals aged 17 or younger, with at least one anxiety disorder, a control group without anxiety or its disorders for comparison, and the use of at least one measure for analyzing cognitive processes. Following the search, a total of 28 studies that met the inclusion criteria were selected. According to the findings across the included studies, the influence of anxiety on cognitive abilities were variable with few statistically significant differences between the group diagnosed with anxiety disorders and the non-anxiety group. For instance, there were relatively few differences regarding processes such as language processing and vocabulary skills, episodic memory or explicit memory of personally experienced events, attention capacities, and problem solving. Nonetheless, individuals with anxiety disorders or heightened anxiety levels performed worse than the non-anxiety groups on tasks requiring spatial and visual working memory or the active processing of incoming sensory information with information in long-term memory, cognitive flexibility or the ability to adjust behaviors and thoughts in different situations. Additional mechanisms analyzed were processing speed, cognitive and inhibitory control, and long term planning.

These mixed findings suggest that cognitive impairments can depend on the type of anxiety disorder a person is diagnosed with as well as age at time of assessment. Additionally, the findings that IQ scores did not exhibit large reductions for those with anxiety disorders indicate that the condition targets specific cognitive abilities rather than a broad spectrum. Despite the mixed results, the review provides some evidence that anxiety disorders can interfere with regulation of behaviors, emotions, and cognitive patterns that are encompassed by the domains of executive functioning and self-regulation. While the review acknowledges the presence of limitations, such as inconsistency and small sample sizes, the authors suggest future directions to strengthen understanding about the relationship between anxiety and cognition and self-regulation. Suggested methods include larger sample sizes, the study of additional executive functioning processes that can provide further insight, and research that can determine whether impairments can serve as a predictor for foundational psychological treatments that can regulate anxiety levels and specific cognitive processes and self-regulatory behaviors (Rabner, 2024).

Cognitive Behavioral Therapy

Cognitive Behavioral Therapy (CBT) is one of the most prominent psychological treatments that utilizes several, smaller interventions to minimize the effects of several psychological disorders and mental states. Extensive research has found convincing evidence that CBT can be influential as the technique heavily focusing on the core principles of cognition, such as thought reconstruction and the reduction of maladaptive behaviors, teaching individuals to form positive perceptions of themselves and the environment, as well as maintain cognitive flexibility in the midst of changing circumstances. While CBT highlights the importance of executive functioning, the approach also focuses on self-regulatory behaviors by exposing individuals to anxiety-arousing stimuli, gradually strengthening controlled emotional responses and managing impulsive responses. A review of meta-analyses was conducted by Hofmann and colleagues in 2013 to analyze efficiency of CBT across a spectrum of psychosocial disorders and mental states. The keywords “meta-analysis”, “efficacy”, “CBT”, “quantitative review”, and “comprehensive review” made up the inclusion criteria, while the exclusion criteria included studies that didn’t include the keywords, as well as studies that were before 2000. From this, a representative sample of 106 meta-analyses were collected. The meta-analyses were categorized for organizational purposes (Hofmann, 2013).

Across many meta-analyses findings, the effect sizes of CBT varied from low to medium for a wide variety of psychological disorders and mental states, such as substance use disorder, schizophrenia, depression, bipolar disorder, somatoform disorder or heavy focus on physical symptoms, eating disorders, insomnia, personality disorders, and antisocial behaviors. Compared to control groups, the participants exhibited positive cognitive and behavioral outcomes when exposed to CBT, but the largest influence was demonstrated on individuals with anxiety disorders and general forms of stress. Specifically, response rates and symptom reduction were higher than control groups and exhibited consistency across studies that implemented follow-up’s, which strengthens evidence about the impact of CBT’s provision of coping

strategies and task-based skillsets that can enable subjects to implement them in their daily lives beyond the treatment period. Overall, the conduction of empirical research enabled the researchers to strengthen the reliability of their conclusions that CBT can be utilized for a variety of conditions, but is highly impactful in reducing the symptoms of stress and various anxiety disorders, as well as strengthening self-regulation and the underlying mechanisms of executive functioning. Additionally, the findings demonstrate that CBT can be applicable to people diagnosed with the most common anxiety disorders today and an effective approach for clinicians and psychologists (Hofmann, 2013).

A scoping review conducted by Dominguez-Pérez and colleagues in 2018 cast emphasis on the neurological mechanisms of CBT's reduction of cognitive, behavioral, and emotional disturbances that characterize anxiety and its related disorders. Moreover, the review was intended to analyze changes and activity in brain systems for individuals receiving CBT, as well as to put forth the dual-process model that informed their hypothesis. Specifically, the dual-process model is a conceptual framework that explains CBT's positive balance of an "impulsive" route, driven by low activation of the limbic system structures responsible for emotional responses and a "reflective" route, driven by high activation of the prefrontal cortex that enables long-term planning, judgement, and especially emotional-regulation. Prior to their review, the researchers followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews which is a checklist that explains the reasoning and procedure for a review or meta-analysis. The inclusion criteria included studies that were published in between 2018 and 2024, studies that included an adult sample with anxiety disorders or general forms of anxiety, studies that were experimental, implemented CBT, and resulted in outcomes that indicated improvement, as well as studies that utilized Functional Magnetic Resonance Imaging (fMRI) or a brain imaging technique that monitors blood flow and, therefore, brain activity by comparing successive MRI scans. The exclusion criteria included studies that were non-experimental, studies where anxiety disorders were not the primary diagnosed disorder, studies that lacked randomization and had a sample size of 15 individuals or less, and studies that did not use fMRI or use the neuroimaging technique in testing the validity of the dual-process model. All the included studies were required to have been peer-reviewed. After checking the inclusion and exclusion criteria for 803 studies from scholarly databases, a final sample of 12 studies was selected. The search terms used included "cognitive behavioral therapy", "anxiety disorders", "specific phobia", and "dual-route model" (Dominguez-Pérez, 2025).

Across all the included studies, general anxiety was reduced for individuals who undertook CBT compared to control groups. Additionally, the results of the included studies displayed improvements in thought and emotion processing among participants, which outlines the primary approaches of CBT in restructuring cognitive processes and strengthening neural networks that enable effortful control of attentional capacities and adapting behaviors to changing circumstances. An important finding identified in this review was that activation of the amygdala, posterior cingulate cortex, and other limbic system structures, which play a role in

generating impulses and emotional reactions, were reduced for individuals diagnosed with Social Anxiety Disorder and Panic Disorder. The heightened activation of prefrontal cortex areas similarly underlines a key improvement in self-regulatory processes that facilitate the cognitive processes of executive functioning. Across the 12 studies, CBT was shown to be effective in enhancing the interconnection between cognition, emotion, and brain regions that not only reduce anxiety symptoms, but also maintain balanced emotional reactivity and adaptive, goal-oriented behaviors necessary for everyday life. One limitation that was acknowledged was that the influence of CBT initiates neuroplasticity or the brain's ability to strengthen and expand neural networks in response to the environment. Also, the researchers stated that there may be other brain regions not specified by the dual-route model that may be influenced by CBT such as neural activity in the precuneus which is responsible for self-referential processes and supporting episodic and spatial memory, and maintaining awareness. This highlights the need for further research about additional sets of brain regions, cognitive processes, and emotional processes that could potentially reflect the effects of CBT. Nonetheless, CBT's neural influences demonstrate evidence for the support of emotional regulation, impulse control, and the cognitive abilities that make up executive functioning (Dominguez-Pérez, 2025).

Mindfulness-Based Interventions

In contrast to CBT, mindfulness-based interventions implement practices such as mindfulness meditation, yoga, deep breathing, awareness of bodily sensations such as heart rate and muscle tension, and focus on the present moment. A combination of these interventions can be beneficial in regulating emotions, behaviors, and high anxiety levels, but researchers have considered whether the interventions influence the cognitive mechanisms of individuals, which includes working memory, goal-oriented behaviors, attentional control, and long-term planning. A quasi-experimental longitudinal study conducted by Cary and colleagues in 2023 attempted to determine whether a Mindfulness-Based Stress Reduction (MBSR) program, or one of the main therapeutic mindfulness approaches, could reduce anxiety among college students. Typically, the duration of MBSR is 8 weeks with daily sessions running for around 2.5 hours and encompassing several mindfulness interventions. While the researchers acknowledged that MBSR can reduce anxiety, support cognitive and self-regulatory behaviors, and enhance psychological well-being, they sought to check for its validity when applied in academic contexts, given that there has been a gradual rise in self-reported anxiety among college students. After obtaining IRB approval and informed consent the researchers assessed MBSR efficacy in the academic context by selecting 34 students who had completed a MBSR curriculum and assigned an additional 110 students to the control group where they would receive a lecture about addiction and the factors behind substance use disorders. The measures used were the 40-item State Trait Anxiety Inventory or a self-report questionnaire that measured anxiety levels in the present moment, the 31-item Short-Self Regulation Questionnaire that measured cognitive self-regulation in terms of planning, impulse control, attention regulation, progress-monitoring, and self-awareness, and the 24-item Five Facet Mindfulness Questionnaire that measured

observance and actions with awareness, non-reactivity, and non-judging. The researchers controlled for confounding variables such as age in their analyses and compared the effects of the interventions on anxiety, mindfulness, and cognitive-regulation between the MBSR group and control group using measures called Analysis of Covariance. The variables were measured before the treatment period, immediately after the 8-week treatment period, and during a follow-up one month after the program concluded (Cary, 2023).

The results obtained highlight the fact that improvement in mindfulness and cognitive self-regulation for students who were in the MBSR group was statistically significant over time compared to the control group. Also, students in the experimental group gradually experienced low state anxiety or anxiety in the present moment compared to the control group. The findings provide convincing evidence that MBSR can facilitate increased mindfulness and self-regulation, which support the cognitive and emotional systems that simultaneously work to minimize anxious behaviors and thoughts in the long run. Moreover, the findings introduce another perspective for MBSR models where mindfulness and cognitive self-regulation are viewed as the central targets of the program. Ultimately the results strengthen connections with additional mechanisms that allow individuals to better regulate anxiety and maintain stable psychological well-being in the midst of changing stressors and environments. Although the researchers acknowledged that their study is potentially limited due to a lack of random assignment and a female majority within the small MBSR group, the findings are generalizable to college students and are beneficial for several positive attentional, behavioral, self-regulatory, mindfulness, and academic outcomes (Cary, 2023).

Although pharmacological treatments are a common approach in managing anxiety symptoms, some individuals do not respond well to prescribed medication. From this, a randomized clinical trial was performed by Hoge and colleagues in 2022 to determine whether MBSR can be as effective as a selective serotonin reuptake inhibitor pharmaceutical, escitalopram, or a common antidepressant medication that is prescribed to treat anxiety disorders by blocking the reuptake of serotonin, in reducing anxiety levels for adults diagnosed with forms of anxiety disorder. After 430 adults were examined by clinicians, a representative sample of 276 adults aged 18-75 with a primary diagnosis of any form of anxiety disorder were collected. Participants who were diagnosed with non-anxiety disorders, had previous experiences with MBSR, and took additional medications were excluded from the study. The participants were randomly assigned to one of two treatments: MBSR or taking Escitalopram in order to reduce differences between the groups. Also, a single-blind procedure was implemented where the subjects did not know which treatment they were receiving. The participants in the MBSR group took part in 2.5 hour group sessions led by a professional instructor who taught mindfulness meditation practices, yoga, awareness of body sensations, and calm regulation of anxiety-arousing emotions and thoughts for a total of 8 weeks. The participants in the medication group took 10 mg of Escitalopram. The amount of Escitalopram given was adjusted at 5 points within the 8 week period. Anxiety symptoms and their intensity were measured using the Clinical Global Impression of Severity scale before the interventions were given, during week 4,

the conclusion of the 8-week treatment period, and during a follow-up 4 and 16 weeks later (Hoge, 2022).

According to statistical analysis, there was convincing evidence that there was no significant difference in effectiveness between Escitalopram and MBSR. In addition, the subjects in both groups displayed similar levels of improvement in anxiety symptoms, cognitive self-regulation, and mindfulness in the weeks following the treatment, indicating the presence of long-term benefits. Furthermore, this suggests that MBSR can serve as an alternative and accessible approach in managing anxiety disorders for individuals with anxiety that have low preference for medicative treatments. Also, an important finding was that subjects in the MBSR group reported slightly fewer adverse side effects compared to those who took Escitalopram, which strengthens the low-risk factors of mindfulness-based interventions. Specifically, this finding may indicate that MBSR can simultaneously support behavioral control, adaptation to distress, thorough self-awareness, and positive cognitive outcomes with minimal impact on physiological well-being. Overall, both MBSR Reduction and Escitalopram are comparably effective in reducing anxiety symptoms, but MBSR further extends its influence by enhancing cognitive and emotional capacities (Hoge, 2022).

Comparing Cognitive Behavioral Therapy and Mindfulness-Based Interventions

A randomized clinical trial was conducted by Goldin and colleagues to determine whether Cognitive Behavioral Group Therapy (CBGT) and MBSR yielded comparable influence on cognitive functioning and emotional regulation in adults diagnosed with Social Anxiety Disorder. Social Anxiety Disorder is one of the most common anxiety disorders with a variety of symptoms revolving around social interaction such as persistent fear of judgement, decreased academic performance, and disruption of daily cognitive functions and thought processes. Additionally, evidence from research studies suggest that CBGT enhances neural activity in the regions of the prefrontal cortex for cognitive and impulse regulation, while MBSR balances emotional reactivity in specific brain regions. Therefore, the researchers sought to address shortcomings in comparison between CBGT and MBSR and gain further insight on any similar or differential influences. The researchers recruited 108 adults with a Social Anxiety Disorder diagnosis and no other mental health conditions. The participants were randomly and equally assigned into one of three experimental groups: 12 weeks of CBT, 12 weeks of MBSR, or the waitlist group where no treatments were given. Participants completed self report measures of anxiety and additional emotions when performing regulatory tasks in a devised social situation. Participants completed fMRIs at three timepoints to measure blood oxygen level-dependent (BOLD) signals or brain activity in regions that manage attention and cognitive reappraisal or the reinterpretation of thoughts in order to minimize anxiety. Approved clinical psychologists and an MBSR professional were present to administer 2.5-hour sessions of CBGT and MBSR, respectively, over the 12 week intervention period (Goldin, 2021).

By examining the fMRI images, the researchers found an increase in BOLD signal activity in the brain areas managing emotional reactivity, acceptance, and cognitive reappraisal

for the CBGT and MBSR groups compared to the waitlist group, indicating that both interventions demonstrate some influence on the analyzed cognitive and behavioral domains. Additionally, models from the study highlighted that reductions in negative emotions were roughly similar for both CBGT and MBSR. The fMRI scans revealed that CBGT generated evident changes in the neural circuits linked to cognitive reappraisal, increased brain activity in the prefrontal cortex, and lessened Social Anxiety Disorder symptoms a year after the treatment period. MBSR was associated with heightened neural activity in regions involved in emotional regulation one year afterward. Nonetheless, the researchers concluded that both interventions did not highlight the presence of specificity as the influences of CBGT and MBSR were not distinguishable from each other when reducing negative emotional reactivity and enhancing acceptance and cognitive reappraisal. Apart from their similarities, the interventions displayed different mechanisms in which CBGT is rooted in direct and effortful alteration of cognitive patterns and thought restructuring and MBSR is rooted in automatic regulation of emotion, attention, and fostering a sense of acceptance. Although the researchers acknowledged that the generalizability of their findings could be increased through the inclusion of additional neural mechanisms, CBGT and MBSR demonstrated the capacity for modulation of anxiety symptoms, evidence of long-term improvement and application, and especially beneficial cognitive, behavioral, and emotional outcomes necessary for navigating engaging environments (Goldin, 2021).

A systematic review and meta-analysis of randomized control trials was conducted by Li and colleagues with a focus on comparing the influences of CBT and mindfulness-based interventions on individuals with anxiety symptoms. The researchers pointed out several key findings such as MBSR and mindfulness-based cognitive therapy are applicable for several psychological disorders and mental health conditions, especially anxiety and its associated disorders. However, a comparison between the two treatments has not been a highly investigated topic, thus informing the researchers' goal on analyzing the effects of the treatments and noting any similarities, differences, and potential long-term changes. Several scholarly databases were utilized to conduct the search with the inclusion criteria encompassing the following: 1) Studies that included sample of individuals with anxiety symptoms or a diagnosis for any anxiety disorder, 2) Studies that were structured as randomized control trials, 3) Studies that included the use of any form of mindfulness-based interventions and CBT for a control group, and 4) Studies that compiled data and results measuring anxiety levels. The exclusion criteria were studies that were not randomized control trials, studies without comparison between the two treatments and measurements of anxiety, and studies that didn't include the desired sample characteristics. The researchers identified 4095 studies that were analyzed based on their respective titles and content, ultimately selecting 11 studies that were selected for the review. Combining the studies, the total number of participants was 819 with a majority of the studies' sample sizes including a higher percentage of women and a roughly equal number of participants with anxiety disorder or anxiety symptoms (Li, 2021).

Across the included studies, there was not a statistically significant difference between the effectiveness CBT and mindfulness-based interventions for reducing anxiety symptoms, suggesting that both treatments exhibit similar influences on the reduction of anxiety scores, as calculated by using the Leibowitz Social Anxiety Scale and the Social Phobia Inventory. Moreover, the results of the included studies indicated that CBT and mindfulness-based interventions further facilitate lessened anxiety by fostering additional processes such as positive thought formation, attentional control, self-awareness, improved cognitive patterns, and regulation of emotions. Additionally, 5 of the 11 studies included a follow-up measurement weeks after experimentation and highlighted consistency in anxiety scores and indications of the long-term effects of CBT and mindfulness-based interventions. Although further research is needed to analyze variability and compare the mechanisms behind CBT and mindfulness-based interventions, the observed similarities across the two treatments demonstrate that anxiety may be influenced by a wider range of different, yet effective cognitive and behavioral regulatory processes (Li, 2021).

Clinical Implications

The findings from the literature review provide a broad description of how CBT and mindfulness-based interventions facilitate the cognitive and self-regulatory processes for those with anxiety. Moreover, they can be applied in medicative practices in which clinicians and health psychologists can confidently implement these therapeutic approaches to navigating anxiety complications for certain individuals. Specifically, both CBT and mindfulness-based interventions, when compared, produce roughly equal reductions of anxiety symptoms even though they operate through different mechanisms. Specifically, CBT focuses on restructuring thought formation, selective attention, the goal-oriented cognitive abilities of executive functioning, while mindfulness-based interventions emphasize self-acceptance, awareness, creating a comfortable environment to navigate, and managing emotional reactivity. In general, CBT is currently viewed as the primary treatment for high levels of anxiety, but mindfulness-based interventions such as MBSR can serve as a viable alternative for patients who do not respond well to structured approaches in order to reduce anxiety symptoms. The notable finding that there was not a significant difference between MBSR and Escitalopram on decreases of anxiety highlights the capacity of mindfulness-based interventions in producing effects in a similar manner to widely-used pharmacological treatments. Nonetheless, the evidence from the literature review provides strong evidence that both CBT and mindfulness-based interventions can be collectively used in addressing the cognitive, behavioral, and emotional mechanisms behind anxiety, thus implementing a more extensive approach to a condition that affects millions worldwide.

Limitations

There are several limitations to the literature review that can be considered when interpreting findings. For example, some of the studies included a sample of participants who

varied in terms of diagnoses, which reduces the generalizability of the findings to larger populations. Another limitation is the variation in the measurement tools used by researchers treating the mindfulness-based intervention groups and CBT groups, which included neuroimaging, certain mindfulness practices such as MBSR, and objective measures such as performance of cognitive and subjective measures such as self-report on anxiety symptoms. This can make it difficult to directly compare results across studies and deduce conclusive findings about which cognitive, behavioral, or emotional outcomes were strongly influenced by which types of treatments. Additionally, only a few studies utilized longitudinal study designs and measured long-term anxiety levels and cognitive and regulatory processes which can make it difficult to conclude whether the effects of CBT and mindfulness-based interventions are sustainable beyond the treatment periods and influential in day-to-day life. Lastly, the findings from each study are potentially limited due to conscious and subconscious bias as a participant's expectations of getting better may influence participant responding and task performance.

Future Directions

Future research should continue to assess CBT and mindfulness-based interventions using similar measures to analyze the mechanisms of executive functioning and self-regulation for individuals with anxiety. Although each treatment consists of different approaches, a combination in the clinical setting can allow psychologists and clinicians to develop new models that provide valuable insight into treatment preferences for individuals, thus prioritizing patient satisfaction alongside alleviation of anxiety symptoms and enhanced behavioral and cognitive outcomes. This can be further strengthened through the use of longitudinal studies to determine if there are promising associations that suggest that the treatment results are long-lasting. Lastly, future research should consider the application of neuroimaging techniques to analyze neural activity and which parts of the brains are active amongst individuals receiving CBT or mindfulness-based interventions. This similarly provides a structured approach to addressing individual differences and understanding the right measures to utilize.

Conclusion

Overall, the purpose of the literature review was to determine whether CBT and mindfulness-based interventions can support executive functioning and self-regulation for individuals with anxiety. Based on the results of meta-analyses and reviews, and experiments, both treatments produced similar effects in improving anxiety symptoms, as well as major cognitive, behavioral, and emotional processes such as maintenance of goal-oriented behavior, reduction of maladaptive thought patterns, emotional regulation, self-awareness, and more. CBT focuses on management of behavioral regulation and refinement of cognitive abilities, while mindfulness-based interventions emphasize an alternative perspective that focuses on attention toward bodily responses, emotions, and internal cognition. Nonetheless, both treatments have the ability to manage self-regulatory behaviors and the goal-oriented mechanisms of executive functioning, thereby easing anxiety levels in the process. As further research seeks to extend the

reviews of the benefits of both interventions, comprehension of the cognitive, behavioral, and emotional aspects of anxiety provides an informed approach to controlling a condition that has affected countless people worldwide. In the process, CBT and mindfulness-based interventions support the psychological well-being of individuals and lay the groundwork for today's continuous advancements in health care.

Works Cited

- “Anxiety Disorders.” World Health Organization, www.who.int/news-room/fact-sheets/detail/anxiety-disorders#:~:text=Key%20facts,and%20school%20or%20working%20life. Accessed 19 Apr. 2026.
- professional, Cleveland Clinic medical. “Executive Function: Skills for Life.” Cleveland Clinic, 22 Dec. 2025, my.clevelandclinic.org/health/articles/executive-function. Accessed 19 Apr. 2026.
- American Psychological Association, www.apa.org/topics/mindfulness/meditation. Accessed 19 Apr. 2026.
- Rabner, Jonathan C., et al. “Cognitive Functioning in Youth with Anxiety Disorders: A Systematic Review - Clinical Child and Family Psychology Review.” SpringerLink, Springer US, 3 June 2024, link.springer.com/article/10.1007/s10567-024-00480-9. Accessed 19 Apr. 2026.
- Hofmann, Stefan G, et al. “The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-Analyses.” Cognitive Therapy and Research, U.S. National Library of Medicine, 1 Oct. 2012, [pmc.ncbi.nlm.nih.gov/articles/PMC3584580/](https://pubmed.ncbi.nlm.nih.gov/articles/PMC3584580/). Accessed 20 Apr. 2026.
- Domínguez-Pérez, Judith, et al. “Neural Mechanisms of Cognitive Behavioral Therapy Efficacy in Anxiety Disorders: A Scoping Review of Fmri-Based Studies That Tested the Dual Model.” MDPI, Multidisciplinary Digital Publishing Institute, 18 Mar. 2025, www.mdpi.com/2075-1729/15/3/493. Accessed 20 Apr. 2026.
- Cary, E. L., et al. (PDF) Self-Regulation Mediates Effects of Adapted Mindfulness-Based Stress Reduction on Anxiety among College Students, 2 May 2023, www.researchgate.net/publication/370470052_Self-regulation_mediates_effects_of_adapted_mindfulness-based_stress_reduction_on_anxiety_among_college_students. Accessed 20 Apr. 2026.
- Hoge, Elizabeth, et al. Mindfulness-Based Stress Reduction vs Escitalopram for the Treatment of Adults with Anxiety Disorders: A Randomized Clinical Trial | Trials | JAMA Psychiatry | Jama Network, 9 Nov. 2022, jamanetwork.com/journals/jamapsychiatry/fullarticle/2798510. Accessed 20 Apr. 2026.
- Goldin, Philippe, et al. Evaluation of Cognitive Behavioral Therapy vs Mindfulness Meditation in Brain Changes during Reappraisal and Acceptance among Patients with Social Anxiety Disorder: A Randomized Clinical Trial | Complementary and Alternative Medicine | Jama Psychiatry | Jama Network, 21 July 2021, jamanetwork.com/journals/jamapsychiatry/fullarticle/2782159. Accessed 21 Apr. 2026.
- Li, Jingjing, et al. “Mindfulness-Based Therapy versus Cognitive Behavioral Therapy for People with Anxiety Symptoms: A Systematic Review and Meta-Analysis of Random Controlled Trials.” Annals of Palliative Medicine, AME Publishing Company, 29 June 2021, apm.amegroups.org/article/view/73159/html. Accessed 20 Apr. 2026.

Etiology, Consequences, and Alleviation Methods of Pediatric Dental Anxiety

By Eliana Park

Abstract

Dental anxiety is a common psychological condition among pediatric patients that may act as a strong deterrent for visiting the dentist. Dental anxiety may hinder oral healthcare, further exacerbating the anxiety as the reluctance and fear of treatment inflates. Such unaddressed dental anxiety among pediatric patients often lasts into adulthood, potentially influencing the anxiety levels of the next generation. Deteriorated oral health across generations implicates broader threats to public health, as oral hygiene acts as an important indicator of systemic health. This study explores the core of dental anxiety through a comprehensive literature review of the etiology, consequences, and alleviation methods. By identifying high-risk groups, this paper aims to recommend interventions to break the intergenerational cycle and enhance public health. This paper hypothesizes that those at greatest risk for experiencing high dental anxiety will be younger, have lower socioeconomic status, and have a family history of dental anxiety. The hypothesis is partially supported in the current body of literature as pediatric patients with lower socioeconomic status are at the greatest risk of exhibiting dental anxiety. However, findings regarding individual factors such as age, gender, and parental influence remain inconsistent, highlighting the need for further research to precisely target high-risk groups. Ultimately, the research indicates the need for effectively targeted interventions for confirmed vulnerable groups, while underlining the importance of future research to clarify debated risk factors.

Introduction

Oral and systemic health have a bidirectional relationship. Accordingly, many systemic diseases (e.g. immune deficiency disorders) appear first in the mouth, and numerous chronic medical conditions (e.g. cardiovascular/pulmonary diseases) are caused by or associated with infectious oral diseases, such as periodontal disease (Yap, 2018). Thus, oral diseases itself could negatively impact quality of life for many populations worldwide, as it is capable of significantly impacting overall health (Fadila, 2024; Yap, 2018). Similar results have been identified for hypertensive patients, where the risk of cardiovascular diseases were increased along with multiple dental caries (Kim et al., 2022), further suggesting the significant effect of oral hygiene on the overall body system. Thus, maintaining good oral health through, for example, brushing, flossing and going to the dentist frequently is not just beneficial for oral health but for the overall systemic health of patients.

The literature overwhelmingly suggests that regularly visiting the dentist is associated with optimal oral health. One study of Saudi Arabian children aged 6-12 years old found that receiving frequent, regular checkups (i.e., visiting the dentist at least once every six months) was associated with significantly less dental pain over six month follow-up (Alhumaid et al., 2018). In contrast, never visiting a dentist or visiting only when in pain was associated with

experiencing persistent dental pain in the past 6 months (Alhumaid et al., 2018). Such trends among pediatric patients indicate the adversely affected quality of life pertaining to pain, which could lead to dental anxiety lasting into adulthood, potentially putting them at risk for more serious health problems stemming from poor oral health. In addition, other researchers identified that studies consistently reported that regular dental attendance was associated with having less dental caries experience, fewer missing teeth and better oral health-related quality of life (Khairuddin et al., 2024). However, inconsistent results were observed for decayed teeth, and no association was found for periodontal conditions (Khairuddin et al., 2024). Despite these inconsistencies, the literature overwhelmingly demonstrates that frequent, regular dentist checkups improve the oral health of children.

Pediatric dental anxiety may act as a hindrance to adhering to regular dental visit recommendations. The associated neglect of oral healthcare among adults is already a grave issue, as indicated in the research using National Study of Adult Oral Health (NSAOH), obtaining data from Australians aged 18+ years. The findings may suggest that dental anxiety may contribute to poor dental attendance among Australian adults (Mejia et al., 2022). Moreover, another paper stated that patients with high dental fear avoid dental visits by postponing or cancelling, which has a negative influence on oral health due to fearful patients completely avoiding dental care (Alenzi et al., 2022). A possible solution to dental anxiety in pediatric patients could be more frequent exposure to lessen anxiety (i.e., resuming frequent dental visits); however, whether this might reduce or worsen dental anxiety might depend on the patient's baseline anxiety level (Gaffar et al., 2014). For higher anxiety patients, it is possible that more frequent dental visits, in the absence of other interventions, could worsen anxiety. For these reasons, children's baseline dental anxiety should be surveyed to determine the most effective targeted interventions to reduce anxiety.

Children's fear in a medical or dental setting could elicit emotional responses including embarrassment, anger and concerns that others were judging them negatively because of their fear (Noble et al., 2020). Such emotional reaction is often accompanied by uncharacteristic behavior such as pushing away the doctor's arm (Noble et al., 2020) and uncooperativeness such as refusing to open their mouth for oral examination (Gao et al., 2013). Similarly, a qualitative study that utilized YouTube as a platform to gather candid, in-depth testimonials and reports stated that the impact of dental anxiety included immediate physical responses such as crying, screaming, and shivering uncontrollably from children and adolescents (Gao et al., 2013). Other responses included psychological effects, such as panic, insecurity and a sense of helplessness particularly prior to the treatment, such as wisdom teeth removal among adolescents (Gao et al., 2013). Another major impact is the clinical consequences of treatment avoidance associated with dental anxiety (Moore et al., 2013), along with the existence of 'vicious cycle dynamics' (Crego et al., 2014), whereby the oral health status of patients declines due to irregular frequencies of dentist visits, often resulting in more painful treatments and thus intensified dental anxiety (Armfield, 2012). Existing literature has established the significant, long-term detriments of pediatric dental anxiety on individual overall health and public health. However, the burden of

the vicious cycle of dental anxiety is not equally distributed across the pediatric population, for example, as prevalence primarily varies across age and caries experience (Grisolia et al., 2021). While the importance of oral healthcare is recognized, the current body of scholarship requires a pivot to identifying the most demographically vulnerable groups to effectively mitigate the overall systemic impact of dental anxiety on public health, enabling the development of a targeted, system level preventative strategies rather than relying on reactive treatment alone.

Role of Individual-level Factors in Pediatric Dental Anxiety

Individual-level factors such as age, gender, previous negative experiences, and infrequent dental visits may greatly contribute to dental anxiety and fear. Dental anxiety decreases in prevalence with age and is more frequently seen in girls (Shim et al., 2015). However, some researchers state that no statistical differences were observed in anxiety levels with respect to patient gender (Mendoza-Mendoza et al., 2015). Such discrepancies may be due to differences in the statistical analysis methods. For example, at the bivariate correlation level, some researchers found a statistically significant association between female gender and dental anxiety in children, but in the regression analysis, the child's gender was no longer a statistically significant predictor, suggesting that this association might have been attributable to another unmeasured factor, such as, the transmission of parental anxiety (Petrović et al., 2024). As the literature surrounding gender remains contradictory, gender may not be an effective indicator for identifying vulnerable populations. Instead, individual factors more universally agreed upon such as previous dental experiences may offer a clearer target for intervention. For instance, previous studies observed that mean anxiety scores were higher in children who had never gone to the dentist. Such results may be due to the fact that a lack of previous experience can generate incorrect thoughts regarding dental procedures (Amorim et al., 2022). Further, previous negative treatment experiences such as injections, drilling and extraction can elicit dental anxiety and significantly negative emotional load (Shim et al., 2015). This may lead to a vicious cycle of dental fear as such anxiety could spread among social circles, resulting in fewer dental visits and deteriorating oral health and so dental fear and anxiety (Shindova et al., 2021). Irregular dental attendance similarly increases the experience of pain and more treatment needs and thus indicates it as a risk factor for the child's dental anxiety (Shindova et al., 2021).

Role of Biological Factors in Pediatric Dental Anxiety

Biological factors of pediatric patients different to that of adult patients may also contribute to intensifying dental anxiety among children. Dental anxiety may be triggered due to younger patients feeling more pain than adults, with several studies finding that younger children were more sensitive to noxious stimuli than older children (Lazin et al., 2024; Tong et al., 2022; El Tumi et al., 2017). The current literature has found significant data involving the activity of medial prefrontal cortex and the medial orbitofrontal cortex pertaining to pain stimuli (Lazin et al., 2024; Tong et al., 2022; El Tumi et al., 2017). Particularly, within 244 participants aged 8–18 years, greater age predicted higher pain tolerance and lower pain intensity and pain

unpleasantness rating to blunt pressure and heat stimuli (El Tumi et al., 2017). Given that pediatric patients have a lower tolerance to blunt pressure (El Tumi et al., 2017), this finding may be applied to the dental setting where tooth extraction or the administration of local anesthesia involves a significant amount of pressure in the gums, potentially provoking intense pain stimuli and subsequent dental anxiety in this demographic. Furthermore, in another study, adolescents exhibited greater pain-evoked neural responses than adults; they showed significant pain-evoked activation in medial prefrontal cortex and deactivation in the medial orbitofrontal cortex, which were not found in adults (Tong et al., 2022). Adolescents have stronger neurologic pain signature responses during pain stimuli than adults, which was tested by exerting pressure at different intervals to mark the changes in the brain images (Tong et al., 2022). Additionally, the results from the multilevel mediation model showed that in adolescence, greater bilateral activation of the amygdala and parahippocampal gyrus significantly predicted greater pain perception above and beyond the effects of stimulus intensity. In contrast, researchers did not find pain-evoked neural responses in amygdala and parahippocampal gyrus as strong predictors of greater pain perception in adults (Tong et al., 2022). These findings suggest a role for these regions in more strongly contributing to pain perception in adolescents (Tong et al., 2022). This may also be applied to the setting where dental treatment triggers pain stimuli more severely in adolescents, standing as evidence for exacerbated dental anxiety in pediatric patients than adult patients. Another study suggests that teeth sensitivity severity increases with age (Lazin et al., 2024), which contrasts the hypothesis of pediatric patients being more prone to feeling sensitive pain, for example, due to having thinner enamel than adults. Although these findings may seem contradictory, they can be reconciled by distinguishing between central pain processing in pediatric patients and structural contributors for dental sensitivity and pain in adults. This may indicate that dental anxiety in pediatric patients is more often worsened by psychological or other outside, non-structural influence rather than structural differences with secondary teeth.

Family Influence on Pediatric Dental Anxiety

Pediatric dental anxiety could additionally be triggered by family influence such as from parents, siblings, or other family structures. Parents are known to subtly transmit their own fear or anxiety to their children, which may be relevant in the dental setting where children's anxiety levels are exacerbated with their parents' (Petrović et al., 2024). As a result of directly witnessing this parental fear, children tend to make clinically significant overproduction of pain, that is, they expect the treatment to be more painful than reality (Shinde et al., 2017). In addition, another study suggests a strong positive correlation between dental anxiety levels of parents and children (Mendoza-Mendoza et al., 2015). However, conflicting evidence suggesting that parental dental anxiety—and parenting style either authoritative or permissive—are not associated with children's dental anxiety do exist (Wu & Gao, 2018). Such discrepancy may be due to the age differences in the pediatric participants of the study; older children's perception of dental treatment is more influenced by their own experiences, such as painful procedures, rather than their parents or other family members' experiences (Wu & Gao, 2018). This age-based

discrepancy highlights the need for age-specific preventative measures for pediatric dental anxiety to alleviate its public health impact; while interventions for younger children may focus on mitigating familial fear, strategies for adolescence may prioritize structural pain management and trauma-informed care.

Previous experiences of siblings may also worsen dental anxiety in pediatric patients, as researchers identify that children with siblings tend to report more dental anxiety as compared with single children (Muhannad, 2020; Wu & Gao, 2018). Such children also reported higher Children Fear Survey Schedule-Dental Subscale (CFSS-DS) scores, which are questionnaires using Likert scales internationally utilized to assess dental anxiety in patients aged from 4 to 13 (Wu & Gao, 2018). This may be due to the exposure to dental procedures during other sibling's therapy or them observing them displaying anxious behaviour while getting treated (Muhannad et al., 2020). Another attribution for such high prevalence of anxiety may be only children tending to internalize parents' expectations, turn predominantly mature, and develop adult behavior due to interacting with adults more frequently (Wu & Gao, 2018). Such perception may also be challenged as some view only-born children to be moody, dependent and shy which may possibly lead to enhanced dental anxiety in only-born children (Wu & Gao, 2018).

Lastly, family structures may influence dental anxiety among pediatric patients. It was found that children in single-parent families reported less dental anxiety, since children may become more independent and are more likely to grow maturity and resilience, making them cope better with stressful situations such as dental visits (Wu & Gao, 2018). They may also receive more care from other caregiver adults such as relatives, which may play a role in their growth and development. Further studies on female patients showed that living in a single parent home indicated a lower level of dental anxiety, as compared with those in nuclear families (two parents; Wu & Gao, 2018), which suggests that socioeconomic status, along with other family influences, might play a significant role in pediatric dental anxiety.

Role of Socioeconomic Status in Pediatric Dental Anxiety

Furthermore, socioeconomic factors such as types of schools, parental education, family income levels, and parental employment status could influence pediatric dental anxiety. For example, a direct, though weak, relationship has been observed between dental caries (cavities) with the category of the school the child attends among children in the city of Curitiba, with children attending public schools having a higher mean DMF-T (Decayed, Missing, and Filled Teeth) index than those attending private schools (Bruzamolin et al., 2018). A study using a short form of Dental subscale of the Child Fear Survey schedule administered to 81 participants aged 8-13 observed a statistically significant relationship between the type of school the child attended and dental anxiety levels (Folayan et al., 2003). However, a cross sectional study among 120 children aged 7-12 concluded that there is no significant difference observed between the occurrence of anxiety in participants from public or private schools (Amorim et al., 2021). Taken together, these studies suggest that socioeconomic and school-related factors may affect pediatric

dental anxiety, although the strength and direction of this association remain inconsistent across studies.

Another socioeconomic determinant is the parental education level, as many researchers found a direct relationship between parental education years and the frequency of flossing and toothbrushing (Bombert et al., 2020). Decreased self-oral hygiene care among children associated with lower years of education may elicit dental anxiety due to increased oral pain from deteriorated oral health. Some researchers conclude that family income levels are also seen as determinants of oral hygiene and so dental anxiety. According to the Brazilian version of the Dental Subscale of the Children's Fear Survey Schedule (B-CFSS-DS) scores, pediatric patients within low income families had a higher rate of dental anxiety (Amorim et al., 2021), supported by patients with a higher ranking of Brazilian economic status having a lower mean DMF-T rate (Bruzamolin et al., 2018). However, research using the scale of CFSS-DS found that the differences between the dental anxiety and the education levels of the mothers and the family income of the parents were not statistically significant (Uzel et al., 2022). Lastly, according to multiple logistic regression analysis, mother's occupation is related to the number of sextants with periodontal bleeding and oral hygiene status (Bashirian et al., 2018). The OHI-S scores of children whose mothers were employed in the government showed that they were 2.27 times more likely to have poor oral hygiene than those with non-employed mothers (Bashirian et al., 2018). This may be due to the fact that government employed mothers have less time to allocate to oral hygiene care of their children than non-employed or self-employed mothers (Bashirian et al., 2018). Overall, socioeconomic factors may also be determinants of oral hygiene and dental anxiety among pediatric patients, although the extents remain debatable.

Role of Environmental Factors in Pediatric Dental Anxiety

Environmental factors including invasive and non-invasive procedures, stimuli triggering the senses and the actions of dental professionals may also contribute to developing pediatric dental anxiety. Invasive procedures such as drilling and injections are often the most feared and thus dental anxiety-triggering factors, supported by the results of the 15-item CFSS-DS-questionnaire—including Likert scales of the fear of ‘injections’ or ‘having a stranger touch you’—drawn by 21 schools in Estonia (Olak et al., 2013). Non-invasive items that may trigger fear such as the fear of not being able to breathe and the fear of the sound of drilling was lower than the general fear of dentistry measured by the Likert-scale (Olak et al., 2013). However, pediatric patients that participated in a study—in Hong Kong towards 405 recruited children aged 9-13 whose dental anxiety was measured by using the Corah Dental Anxiety Scale (CDAS) and CFSS-DS—related their dental anxiety to drilling and further stimuli (sound and light) associated with drilling (Wu & Gao, 2018). In addition, “having a stranger touch you” also appeared as a major fearful situation for the respondents (Wu & Gao 2018). Although being touched by a stranger may not be an anxiety-triggering factor for adults, it may impose considerable stress on pediatric patients (Wu & Gao, 2018). Similarly, the fear of specific stimuli was identified as the most common source of dental anxiety, with the sight and sensation of the

anaesthetic needle and the sight, sound and sensation of the drill being rated the most fear-eliciting stimuli, using DMF-T Index and Löe and Silness Gingival Index (GI) among 12–15-year-old schoolchildren, in Irbid Governorate/northern Jordan (Taani et al., 2005). Furthermore, when compared to non-anxious patients, anxious patients reported fear towards the smell associated with dental materials, and the sight and sound of instruments provoked during treatments (Shindova et al., 2021). In addition, dentists and dental auxiliaries could also influence the levels of pediatric dental anxiety. A qualitative study using YouTube as its platform to gain testimonials examined that dental professionals' actions such as commanding loudly, sternly and impatiently to pediatric patients triggers dental anxiety (Gao et al., 2013). Professionals' lack of clinical skills may also elicit mixed feelings of panic and anger, contributing to dental anxiety perhaps even in later visits (Gao et al., 2013). Correspondingly, a literature review in 2021 stated that a positive dentist-patient relationship can have a promising effect on pediatric dental anxiety, whereas an unpleasant first visit could influence further treatment and lead to dental anxiety (Shindova et al., 2021). According to the review, several studies indicate that unfriendly dental staff, negative dentist behaviour and shorter mean treatment time are risk factors for DA (Shindova et al., 2021), which offer potential intervention targets for providers treating fearful pediatric patients.

Discussion

The synthesis of available research suggests that dental anxiety is a complex phenomenon impacted by multiple elements interdependently. Despite some contradictory findings within the literature, the evidence generally suggests high-risk profiles of patients that fall into several key categories. Specifically, female gender, higher pain sensitivity determined by younger age, parents' and siblings' existing anxiety levels, and low socioeconomic status are patient-level factors that increase a child's risk for dental anxiety (Shim et al., 2015; Petrović et al., 2024; Muhannad et al., 2020; Brazamolin et al., 2018). Additionally, factors specific to the dental office, such as noise (Wu & Gao, 2018) and anesthesia (Olak et al., 2013) are modifiable environmental factors that might increase or decrease patient anxiety. Based on these findings, this review recommends implementing screening by pediatric dental staff at patient intake that could also help to track anxiety over time and determine the efficaciousness of interventions. Additionally, such screening using dental anxiety scales may identify patients and family members with dental anxiety early on to offer helpful and early in-clinic interventions. Identifying and treating such conditions early on could prevent worsening dental anxiety within the patient as well as the potential social contagion to siblings or, in adulthood, the patient's children.

Potential ways to identify who is at the greatest risk of dental anxiety early on includes conducting a brief survey prior to treatment using dental anxiety scales. For pediatric patients to fill out before treatment, this analysis of the literature suggests that the DFSS-SF remains the most appropriate instrument as it is a simplified version of the 'golden standard' CFSS-DS. While the CFSS-DS yields important information, it may be too time-consuming when rapid

screening is needed, and the 15 items could be answered by parents, resulting in bias. Organizing a dental anxiety scale for the patients' parents may also be adequate as the level of dental anxiety of parents could greatly affect the anxiety levels of their children, making the survey of parents useful in terms of identifying potential dental anxiety of pediatric patients. For parents, either MDAS that could be briefly done in waiting rooms, or DFS, which finds out *why* a patient is anxious, providing a practical way to identify whether the children could be affected the same way as their parents. Although it is more time-consuming than MDAS, this may be negotiable considering that the parents could fill it out during the pediatric patient's procedure. In increasing the number of patients surveyed, understanding the factors associated with pediatric dental anxiety can be improved. Nonetheless, focusing on anxiety reducing strategies for high-risk pediatric patients may improve efficiency and avoid unnecessary intervention—avoiding potential overwhelm—for those who do not exhibit significant dental anxiety, while allowing clinicians to adapt more intensive interventions for patients who are more likely to benefit.

One of the most heavily emphasized methods to reduce pediatric dental anxiety during dental procedures is non-pharmacological behavior management, or behavior guidance, techniques (American Academy of Pediatric Dentistry, 2025). Examples of such methods include 'Tell-Show-Do', where the children could reduce the fear of unexpected stimulation by the dentist's warning, or 'Ask-Tell-Ask', where the dentist asks the patient what they would like to know, and give the information in simple terms, and ask once more to check the patient's understanding, easing children's fear of unknown procedures (AAPD, 2025). Building upon the environmental factors triggering pediatric dental anxiety, appeasing sensory overload could also serve as an effective intervention. A sensory adapted dental environment such as dimmed lights, soothing music, soft blankets etc., could relieve the potential anxiety the children could hold by calming their senses down. Furthermore, when non-pharmacological interventions are insufficient, pharmacological methods such as Nitrous Oxide (Mukundan & Gurunathan, 2023) sedation may be recommended to reduce pain and thus the resultant dental anxiety among pediatric patients.

Conclusion/ Implications

The comparative nature showing both sides of discussion from the existing body of literature may serve as a practical guide for future studies. The findings of who may be at most risk are not definitive as a number of studies indicate that dental anxiety occurs regardless of some factors such as gender, and pain sensitivity in teeth increases with age, the wide age range among participants distorting results and the positive yet low correlation of certain factors such as socioeconomic elements. Furthermore, the generalizability of findings may be limited as the methodologies of research vary greatly ranging from self-report interviews to standard survey methods of anxiety. To conclude, while the primary risk factors are arguably identifiable, the precise acknowledgement of the patient group *most* vulnerable to dental anxiety by analyzing how numerous variables interact to determine the wider public health effects remain areas for future investigation.

Further potential interventions based on the causes of pediatric dental anxiety explored in this paper to improve public health include incorporating quick surveys in the waiting rooms—for both the pediatric patient and their guardians. The patient may be given a short survey asking them to rate ‘how scared they are’ on a scale of 1 to 10, 1 being relaxed and 10 being highly anxious. so the doctors and hygienists can refer to it. For the guardian, questions asking about their own, or other family members’—such as siblings of the patient— experiences of dental anxiety may be relevant as it may have a heavy influence on the patient’s anxiety more so than the patient perceives. Other questions may aim to determine whether the child has visited a dentist before and, if so, whether the patient experienced invasive treatments such as drilling, extraction or injection that might affect their current anxiety levels. Additional queries asking whether the child visits the dentist regularly, or how the guardian would rate their personal oral hygiene or oral care skills may be applicable for dentists or hygienists to swiftly predict whether or not the patient may exhibit dental anxiety prior to the procedure.

In the final analysis, modifications in the behavior of dental staff or changes in the clinic environment could reduce dental anxiety. A careful introduction of the dentist and hygienists to the patient before treatment could ease the fear of “being touched by strangers” among children. Staff members should be ensured to hold a calm, comforting yet professional tone, and try to form a rapport with patients, especially during the first visit. As the sight of needles also greatly elicits fear, the staff could gently recommend patients to close their eyes, since blocking their vision—for example, using eye masks—could make them feel a sense of lacking control over the situation. Earplugs, or headsets with calming music may cover the sound of drills, and citrus, fresh, or woody scents may help mask the odor of dental instruments or eugenol. Overall, interventions managing surveys, staff behavior, and dental environments may enhance public health by enabling speedy dental anxiety predictions to inform staff members and by mitigating existing or potential dental anxiety.

Works Cited

- Alenezi, Ali A., and Hamad S. Aldokhayel. "The Impact of Dental Fear on the Dental Attendance Behaviors: A Retrospective Study." *Journal of Family Medicine and Primary Care*, vol. 11, no. 10, Oct. 2022, pp. 6444–50, https://doi.org/10.4103/JFMPC.JFMPC_1030_22.
- AlHumaid, Jehan, et al. "Dental Visit Patterns and Oral Health Outcomes in Saudi Children." *Saudi Journal of Medicine & Medical Sciences*, vol. 6, no. 2, May 2018, pp. 89–94, https://doi.org/10.4103/SJMMS.SJMMS_103_17.
- American Academy of Pediatric Dentistry. Behavior Guidance for the Pediatric Dental Patient. 2024, https://www.aapd.org/globalassets/media/policies_guidelines/bp_behavguide.pdf.
- Amorim, Camila Silva de, et al. "The Effect of Socioeconomic Aspects and Dental History on Pediatric Patients' Dental Anxiety." *Brazilian Oral Research*, vol. 36, 2022, p. e106, <https://doi.org/10.1590/1807-3107BOR-2022.VOL36.0106>.
- Armfield, Jason M. "What Goes around Comes around: Revisiting the Hypothesized Vicious Cycle of Dental Fear and Avoidance." *Community Dentistry and Oral Epidemiology*, vol. 41, no. 3, June 2013, pp. 279–87, <https://doi.org/10.1111/CDOE.12005>.
- Bashirian, Saeed, et al. "Socio-Demographic Determinants as Predictors of Oral Hygiene Status and Gingivitis in Schoolchildren Aged 7-12 Years Old: A Cross-Sectional Study." *PLOS ONE*, vol. 13, no. 12, Dec. 2018, p. e0208886, <https://doi.org/10.1371/JOURNAL.PONE.0208886>.
- Bombert, Filipa, et al. "Sociodemographic Factors Associated with Oral Health in 12-Year-Old Adolescents: Hygiene Behaviours and Health Appointments. A Cross-Sectional National Study in Portugal." *International Dental Journal*, vol. 68, no. 5, Oct. 2018, pp. 327–35, <https://doi.org/10.1111/idj.12390>.
- Crego, Antonio, et al. "From Public Mental Health to Community Oral Health: The Impact of Dental Anxiety and Fear on Dental Status." *Frontiers in Public Health*, vol. 2, no. FEB, Feb. 2014, p. 81368, <https://doi.org/10.3389/FPUBH.2014.00016/TEXT>.
- El Tumi, H., et al. "Age-Related Changes in Pain Sensitivity in Healthy Humans: A Systematic Review with Meta-Analysis." *European Journal of Pain (London, England)*, vol. 21, no. 6, July 2017, pp. 955–64, <https://doi.org/10.1002/EJP.1011>.
- Fadila, Wisnu, et al. "Family's Socioeconomic and Demographic Factors on Elements of Children's Dental and Oral Health: A Scoping Review." *Clinical and Experimental Dental Research*, 2024, <https://doi.org/10.1002/CRE2.818>.
- Gaffar, Balgis O., et al. The Prevalence, Causes, and Relativity of Dental Anxiety in Adult Patients to Irregular Dental Visits. www.smj.org.sa.
- Gao, Xiaoli, et al. "Dental Fear and Anxiety in Children and Adolescents: Qualitative Study Using Youtube." *Journal of Medical Internet Research*, vol. 15, no. 2, Feb. 2013, p. e2290, <https://doi.org/10.2196/jmir.2290>.

- Grisolia, Barbara Monteiro, et al. "Prevalence of Dental Anxiety in Children and Adolescents Globally: A Systematic Review with Meta-Analyses." *International Journal of Paediatric Dentistry*, vol. 31, no. 2, Mar. 2021, pp. 168–83, <https://doi.org/10.1111/IPD.12712>.
- Kim, Jinkwon, et al. "Association between Oral Health and Cardiovascular Outcomes in Patients with Hypertension: A Nationwide Cohort Study." *Journal of Hypertension*, vol. 40, no. 2, Feb. 2022, pp. 374–81, <https://doi.org/10.1097/HJH.0000000000003022>.
- Lazin, Muhamamd Amirul Mat, et al. "Factors Influencing Tooth Sensitivity: Insights From the Hospital Universiti Sains Malaysia Using Bootstrap-Enhanced Ordinal Regression." *Cureus*, vol. 16, no. 7, July 2024, p. e64641, <https://doi.org/10.7759/CUREUS.64641>.
- Mejia, G. C., et al. "Immigrants Experience Oral Health Care Inequity: Findings from Australia's National Study of Adult Oral Health." *Australian Dental Journal* vol. 68, no. 1, 2023, pp. 7–18. PubMed Central, PMC10952364, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10952364/>.
- Mendoza-Mendoza, Asunción, et al. "Dental Fear in Children: The Role of Previous Negative Dental Experiences." *Clinical Oral Investigations* 2014 19:3, vol. 19, no. 3, Dec. 2014, pp. 745–51, <https://doi.org/10.1007/S00784-014-1380-5>.
- Mohd Khairuddin, Aina Najwa, et al. "Impact of Dental Visiting Patterns on Oral Health: A Systematic Review of Longitudinal Studies." *BDJ Open* 2024 10:1, vol. 10, no. 1, Mar. 2024, pp. 18-, <https://doi.org/10.1038/s41405-024-00195-7>.
- Moore, Rod, et al. "The Contribution of Embarrassment to Phobic Dental Anxiety: A Qualitative Research Study." *BMC Psychiatry* 2004 4:1, vol. 4, no. 1, Apr. 2004, pp. 10-, <https://doi.org/10.1186/1471-244X-4-10>.
- Mukundan, Divya, and Deepa Gurunathan. "Effectiveness of Nitrous Oxide Sedation on Child's Anxiety and Parent Perception During Inferior Alveolar Nerve Block: A Randomized Controlled Trial." *Cureus*, vol. 15, no. 11, Nov. 2023, <https://doi.org/10.7759/CUREUS.48646>.
- Murad, Muhannad H., et al. "Evaluating Factors Associated with Fear and Anxiety to Dental Treatment-A Systematic Review." *Journal of Family Medicine and Primary Care*, vol. 9, no. 9, Sept. 2020, pp. 4530–35, https://doi.org/10.4103/JFMPC.JFMPC_607_20.
- Noble, Fiona, et al. "'I Would Rather Be Having My Leg Cut off Than a Little Needle': A Supplementary Qualitative Analysis of Dentally Anxious Children's Experiences of Needle Fear." *Dentistry Journal* 2020, Vol. 8, vol. 8, no. 2, May 2020, <https://doi.org/10.3390/DJ8020050>.
- Olak, Jana, et al. "Children's Dental Fear in Relation to Dental Health and Parental Dental Fear." PubMed, 2013, <https://sbdmj.lsmuni.lt/131/131-04.pdf>.
- Petrović, Dorotea, et al. "The Role of Family Factors in the Development of Dental Anxiety in Children." *Medicina (Kaunas, Lithuania)*, vol. 60, no. 1, Jan. 2024, <https://doi.org/10.3390/MEDICINA60010180>.
- Reddy Elicherla, Niharika, et al. "Evaluation of the Effectiveness of Tell-Show-Do and Ask-Tell-Ask in the Management of Dental Fear and Anxiety: A Double-Blinded

- Randomized Control Trial.” *Journal of Dental Anesthesia and Pain Medicine*, vol. 24, no. 1, Feb. 2024, pp. 57–65, <https://doi.org/10.17245/JDAPM.2024.24.1.57>.
- Shim, Youn-Soo, et al. “Dental Fear & Anxiety and Dental Pain in Children and Adolescents; a Systemic Review.” *Journal of Dental Anesthesia and Pain Medicine*, vol. 15, no. 2, 2015, p. 53, <https://doi.org/10.17245/JDAPM.2015.15.2.53>.
- Shinde, Swapnali Dilip, and Rahul J. Hegde. “Evaluation of the Influence of Parental Anxiety on Children’s Behavior and Understanding Children’s Dental Anxiety after Sequential Dental Visits.” *Indian Journal of Dental Research : Official Publication of Indian Society for Dental Research*, vol. 28, no. 1, Jan. 2017, pp. 22–26, https://doi.org/10.4103/IJDR.IJDR_181_16.
- Shindova, Maria P., and Ani B. Belcheva. “Dental Fear and Anxiety in Children: A Review of the Environmental Factors.” *Folia Medica*, vol. 63, no. 2, 2021, pp. 177–82, <https://doi.org/10.3897/FOLMED.63.E54763>.
- Taani, D. Q., et al. “Dental Anxiety in Children and Its Relationship to Dental Caries and Gingival Condition.” *International Journal of Dental Hygiene*, vol. 3, no. 2, 2005, pp. 83–87, <https://doi.org/10.1111/J.1601-5037.2005.00127.X>.
- Tong, Han, et al. “Processing of Pain by the Developing Brain: Evidence of Differences between Adolescent and Adult Females.” *Pain*, vol. 163, no. 9, Sept. 2022, pp. 1777–89, <https://doi.org/10.1097/J.PAIN.0000000000002571>.
- Uzel, İlhan, et al. Evaluation of the Risk Factors of Dental Anxiety in Children. 2022, <https://doi.org/10.4274/jpr.galenos.2022.79990>.
- Wu, Lingli, and Xiaoli Gao. “Children’s Dental Fear and Anxiety: Exploring Family Related Factors.” *BMC Oral Health* 2018 18:1, vol. 18, no. 1, June 2018, pp. 100-, <https://doi.org/10.1186/S12903-018-0553-Z>.
- Yap, Adrian. “Oral Health Equals Total Health: A Brief Review.” *Journal of Dentistry Indonesia*, vol. 24, no. 2, Aug. 2017, <https://doi.org/10.14693/jdi.v24i2.1122>.

The Effect of Taxation on Economic, Social, and Political Indicators By Brian Chou

Abstract

The purpose of this study is to investigate how taxation affects economic, social and political indicators for 53 developing and 51 developed countries. Using regression analysis, the results indicate that in developed countries, higher taxes are associated with higher levels of GDP and unemployment, and lower levels of life expectancy, current account balance, rule of law, and government effectiveness. In developing countries, higher taxes lead to higher life expectancy and reduce unemployment. The findings highlight the role of developing countries for the effect of taxation on economic, social and political outcomes of countries. While existing studies focus on the effect of taxation on macroeconomic indicators; this paper extends the literature by incorporating social and political indicators.

Keywords Taxation, macroeconomic indicator, social indicator, political indicator

Introduction

Taxation is one of the most important instruments of fiscal policy, serving as the primary social government revenue, required to finance public goods, infrastructures, and social welfare programs. However the social, economic, and political impact of taxation is an ongoing debate among economists and policy makers. There is some controversial discussion that higher taxes reduce disposable income, on the other hand tax revenue can stimulate long term development. This argument is reasonable when considering that countries are at different developing levels, difference in institutional qualities, and administrative capacities.

Investigating the effect of taxation on economic, social, and political indicators is important for several groups for multiple reasons. For policy makers, understanding how tax structure influences growth, income distribution, and welfare helps them design a more efficient fiscal system. The empirical evidence also helps economists to understand the link between taxation and economic decisions, institutional quality, and development levels of countries. For business and investors, insights into taxation policy provide valuable information about investment, political risk, and sustainability. Finally, for society examining these relationships clarifies how tax policies affect daily life.

Despite the extensive literature on this study, many studies focus on indicators such as GDP growth or FDI inflows on specific regions that remains a need for comprehensive analysis that examines the effect of taxation on socioeconomic indicators. This paper addresses the gap in investigating how taxation influences a wide range of indicators. Specifically, the aim of this paper is to investigate the effect of taxation on economic, social, and political indicators for 53 developing and 51 developed countries. Moreover, this study tests whether the level of economic development moderates the relationship between taxation and the socioeconomic and political indicators. Regression analysis shows that, in developed countries, taxation is positively associated with GDP and unemployment, while it is negatively associated with life expectancy,

current account balance, rule of law, and government effectiveness. In contrast, in developing countries, higher taxes are associated with higher levels of life expectancy and lower levels of unemployment. These findings support empirical evidence that developing levels of countries play a significant role in this relationship.

This paper makes contributions on the ongoing literature in two main ways: first, rather than analyzing all countries as a single group, this study divides the countries into two categories as developing and developed to understand the effect of developing level of countries on these relationships. Most fiscal policy papers focus on macroeconomic indicators such as inflation or trade balance, we expand this view to include social and political indicators such as life expectancy, government effectiveness, and political stability.

This paper proceeds as follows: Section 2 provides background information on the existing literature, Section 3 highlights the empirical findings, and Section 4 concludes the paper.

Literature Review

Various methodologies have been used and employed to investigate the effect of taxation on social, economic, and political indicators such as VAR (Karras and Furceri, 2009), Panel Data Regression (Elshani et al., 2024; Wang, 2024), ARDL (Shabbir et al., 2025; Malik and Meraj, 2024; Bashir et al., 2024; Bozatli and Akca, 2024), Regression Analysis (Oluchi and Abel, 2024; Freund and Gagnon, 2017), Static Panel Data Analysis (Chettri et al., 2023), GMM (Zirgulis and Šarapovas, 2017; Ballesta and Yague, 2024; Bah, 2023), Ordinary Least Square (Atan, 2013; Soto et al., 2024), Linear Mixed Model (Mustafa et al., 2024), Cluster Analysis (Alfo et al., 2023), MF-VAR (Oyeleke and Onatunji, 2024), FGLS (Alshubiri, 2024).

There are many studies that examine the effect of taxation on a wide range of indicators including FDI (Alshubiri, 2024), Government Effectiveness (Bozatli and Akca, 2024), Political Stability (Soto et al., 2024), Rule of Law (Bah, 2023), LEV (Ballesta and Yague, 2024), Climate Change (Bashir et al., 2024), Inflation (Oyeleke and Onatunji, 2024), GDP Growth (Alfo et al., 2023), Corporate Total Asset (Wang, 2024), Polity Index (Malike and Meraj, 2024), Control of Corruption (Mustafa et al., 2024), Real Exchange Rate (Freund and Gagnon, 2017), Gross Private Investment (Atan, 2013), Unemployment (Zirgulis and Šarapovas, 2017), Life Expectancy (Chettri et al., 2023), Trade Openness (Shabbir et al., 2025), Social Security Contribution (Elshani et al., 2024), Profits and Capital Gains (Karras and Furceri, 2009).

Many Studies have focused on various countries and regions such as Central Europe (Alshubiri, 2024), the United States (Elshani et al., 2024), Nigeria (Oluchi and Abel, 2024), Pakistan (Malik and Meraj, 2024), South Asian Countries (Chettri et al., 2023), Advanced Economies (Freund and Gagnon, 2017), China (Wang, 2024), OECD (Alfo et al., 2023), G7 (Bashir et al., 2024), EU (Ballesta and Yague, 2024), Sub-Saharan Africa (Bah, 2023), The Netherlands (Bozatli and Akca, 2024).

Various Studies have investigated the effect of taxation on social, economic, and political development for example, Karras and Furceri (2009) found that higher taxes reduce growth in Europe and Elshani et al. (2024) concluded that taxes can stimulate investment and sustain GDP

growth. Similarly, Onatunji and Oleyeke (2024) found that tax revenues contribute to controlling inflation and improving economic sustainability.

In developing countries, the relationship between taxation and economic outcomes often yields mixed results. For example, Atan (2013) and Olechi and Abel (2024) reported that in Nigeria, tax policy has limited effect on inflation and unemployment but positive effect on FDI inflow. Malike and Meraj (2024) emphasized taxation and political stability are statistically insignificant drivers in Pakistan, meanwhile Chettri et al. (2013) offers a different perspective on South Asia, demonstrating that FDI and social developing outcomes can improve tax collection efficiency. Bozatli and Akca (2024) found that environmental taxes are effective in improving environmental quality.

Table 1. Extensive Summary of Literature on the Effect of Taxation on Socioeconomic and Political Indicator

Author	Year	Variables	Region/Country	Methodology	Key Findings	Time period
Karras & Furceri	2009	Taxes and GDP	Central Europe	VAR	Higher taxes reduce GDP per capita	1965-2003
Elshani et al.	2024	GDP, PIT, CIT, VAT	USA, EU	Panel Regression	PIT and VAT reduce growth, while CIT and SGST have positive effects.	1995-2020
Shabbir et al.	2025	Tax, Inflation, FDI	Pakistan	ARDL	Trade openness and low taxes increase FDI.	1991-2024
Oluchi & Abel	2024	CIT, VAT, FDI	Nigeria	Regression	CIT and VAT boost FDI.	2000-2020
Chettri et al.	2023	FDI, Life Expectancy, Tax, Trade	South Asia Countries	Panel Data	Trade openness and FDI, positively influence tax revenue.	2008-2017
Zirgulis & Šarapovas	2017	Effective Marginal Tax Rate, Unemployment, Working Age Population	Global	GMM	Higher corporate tax rates are linked to increased unemployment.	2001-2012
Atan	2013	Unemployment, Inflation, Investment	Nigeria	OLS	Tax policy has little impact on controlling inflation and unemployment.	1970-2008
Freund & Gagnon	2017	VAT, Exchange Rate, Trade Balance, CAB	34 Advanced Economies	Regression	Consumption on taxes raise exchange rate	1970-2015
Mustafa et al.	2024	Tax Evasion, Government Effectiveness, Corruption, FDI	Global	Mixed Model	Government efficiency reduces tax evasion.	2010-2015
Malik & Meraj	2024	Income Tax, Corporate Tax, Polity Index	Pakistan	ARDL	Taxation and political stability strongly influence economic growth.	1979-2019
Wang	2024	CIT, Total Asset, , Average Wage	China	Panel Data Regression	CTR rises around government bodies and declines in private firms, and high-investment regions.	1998-2009
Alfo et al.	2023	Taxation, GDP Growth	OECD	Panel Data, Cluster Analysis	Tax cuts positively affect real income growth across countries.	1965-2010

Oyeleke & Onatunji	2024	Government Revenue & Expenditure	Nigeria	MF-VAR	Fiscal expansion boosts growth while tax revenue helps control inflation.	2010-2022
Bashir et al.	2024	Climate Change, Taxation, Trade Openness	OECD & G7	ARDL	Environmental taxes, renewable energy, and green technology reduce environmental degradation, while growth and trade worsen it.	1994-2021
Ballesta & Yague	2024	Total Effective Tax Rate, LEV, ROA, Net Sales	EU	GMM	Tax-avoiding SMEs use longer debt maturities and less short-term debt, as tax savings improve their financial structure.	2004-2020
Bah	2023	Taxes, Rule Of Law, Informal Sector	Sub Saharan Africa	GMM	Better institutional quality boosts tax revenue.	1996-2019
Soto et al.	2024	HDI, FDI, Political Stability	Global	OLS	FDI and taxes reduce income inequality, with taxes amplifying this effect.	2000-2021
Bozatli & Akca	2024	Government Effectiveness, Tax	Netherland	ARDL	Environmental policies and renewable energy improve environmental quality while economic growth harms it.	1996-2021
Alshubiri	2024	Revenue, Personal Remittances Received, Real Interest Rate	Central Europe & USA	FGLS, GMM, PMG, ARDL	In developed countries, FDI inflows are linked to lower long-term tax revenue, while in developing countries, FDI boosts long-term tax revenue.	2006-2020

Empirical Findings

Data

The appendings present various macroeconomic indicators and tax variables sourced from World Bank World Development Indicator. The most available data were used in the analysis; for our sample, this corresponds to the year 2023. The variable includes Taxes on goods and services as a percentage of revenue, GDP per capita as an annual percentage of growth, inflation as a consumer prices, foreign direct investment net inflows as a percentage of GDP, labor force participation rate as a percentage of total population ages 15-64, life expectancy, unemployment as a percentage of total labor force, current account balance as a percentage of GDP, government effectiveness, political stability, rule of law, CPIA debt policy rating.

Descriptive statistics

Table 2. Descriptive Statistics for Developing and Developed Countries

Variable		Obs	Mean	Std. Dev.	Min	Max
GDP Growth	Developing	53	3.399	2.843	-3.241	15.218
	Developed	51	.938	2.272	-4.541	5.524
Taxes	Developing	53	35.534	10.199	1.668	51.519
	Developed	51	29.136	9.914	1.899	57.05
Inflation	Developing	51	7.305	4.96	.743	30.219
	Developed	49	6.491	7.426	.235	53.859
Foreign inflows	Developing	52	3.149	3.443	.255	18.491
	Developed	51	2.469	20.036	-71.719	113.224
Labor force	Developing	51	65.036	10.978	34.794	86.342
	Developed	51	74.651	5.719	59.364	86.713
Life expectancy	Developing	53	71.592	4.995	58.816	79.602
	Developed	51	79.854	3.088	73.255	84.056
Unemployment	Developing	21	7.613	5.446	.733	23.381
	Developed	46	5.497	2.278	2.151	12.179
CAB	Developing	38	-2.656	5.982	-19.78	13.574
	Developed	40	1.075	5.924	-11.304	17.688
Gov. effectiveness	Developing	45	-.318	.56	-1.98	.791
	Developed	41	.932	.699	-.712	2.317
Political stability	Developing	45	-.277	.775	-2.378	1.1
	Developed	41	.423	.646	-1.465	1.424
Rule of law	Developing	45	-.375	.617	-2.212	.901
	Developed	41	.832	.786	-1.188	1.967
CPIA debt policy rating	Developing	31	3.476	.743	2	5
	Developed	4	3.418	.58	2.714	4.125

Table 2 presents descriptive statistics for macroeconomic and tax variables used in the analysis for developing and developed countries, respectively. For GDP growth, developing countries (53 observations) exhibit a higher mean of 3.399 with greater variability (SD = 2.843), compared to developed countries (51 observations), which have a lower mean of 0.938 and a standard deviation of 2.272. Tax revenue is also higher on average in developing countries (mean = 35.534, SD = 10.199) than in developed countries (mean = 29.136, SD = 9.914), though both show wide ranges.

Inflation is slightly higher in developing countries (mean = 7.305, SD = 4.96) than in developed countries (mean = 6.491), but developed countries display much greater dispersion (SD = 7.426), with extreme maximum values. Foreign direct investment (FDI) inflows are more stable in developing countries (mean = 3.149, SD = 3.443), whereas developed countries show substantial volatility (mean = 2.469, SD = 20.036), including large negative and positive extremes.

Labor force participation is lower in developing countries (mean = 65.036, SD = 10.978) compared to developed countries (mean = 74.651, SD = 5.719), while life expectancy is also significantly lower in developing countries (mean = 71.592) than in developed countries (mean = 79.854). Unemployment rates are higher and more dispersed in developing countries (mean = 7.613, SD = 5.446) relative to developed countries (mean = 5.497, SD = 2.278).

In terms of external balance, developing countries run an average current account deficit (mean = -2.656), whereas developed countries show a surplus on average (mean = 1.075), though both groups have similar variability. Institutional quality indicators—including government effectiveness, political stability, and rule of law—are consistently lower (and negative on average) in developing countries, while developed countries exhibit positive and higher mean values, reflecting stronger institutions.

Finally, CPIA debt policy scores are similar across both groups, with developing countries (31 observations) having a mean of 3.476 and developed countries (only 4 observations) showing a comparable mean of 3.418, though the small sample size for developed countries limits comparability.

Model

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \mu_i$$

Y_i is the dependent variables for observation i , which refers to GDP Growth, Inflation, FDI Inflows, Labor Force, Life Expectancy, Unemployment, Current Account Balance, Government Effectiveness, Political Stability, Rule of Law, CPIA debt rating

α_0 is the constant term, representing the expected value of the dependent variable when the independent variable is zero.

α_1 is the coefficient for independent variables which is taxes on goods and services. The coefficient shows how much the dependent variable changes when the independent variable changes by 1 unit.

μ_i is the error term, which represents the difference between the actual value and the predicted value from the model.

To investigate the effect of taxation on social, economic, and political indicators, we estimated the following models.

$$GDP\ Growth_i = \alpha_0 + \alpha_1 Taxation_i + \mu_i \text{ (Model 1)}$$

$$Inflation_i = \alpha_0 + \alpha_1 Taxation_i + \mu_i \text{ (Model 2)}$$

$$FDI\ inflows_i = \alpha_0 + \alpha_1 Taxation_i + \mu_i \text{ (Model 3)}$$

$$Labor\ force_i = \alpha_0 + \alpha_1 Taxation_i + \mu_i \text{ (Model 4)}$$

$$Political\ Stability_i = \alpha_0 + \alpha_1 Taxation_i + \mu_i \text{ (Model 5)}$$

$$Life\ expectancy_i = \alpha_0 + \alpha_1 Taxation_i + \mu_i \text{ (Model 6)}$$

$$Unemployment_i = \alpha_0 + \alpha_1 Taxation_i + \mu_i \text{ (Model 7)}$$

$$\begin{aligned} \text{Current Account Balance}_i &= \alpha_0 + \alpha_1 \text{Taxation}_i + \mu_i \text{ (Model 8)} \\ \text{Government Effectiveness}_i &= \alpha_0 + \alpha_1 \text{Taxation}_i + \mu_i \text{ (Model 9)} \\ \text{Rule of Law}_i &= \alpha_0 + \alpha_1 \text{Taxation}_i + \mu_i \text{ (Model 10)} \end{aligned}$$

Findings

Figure 1. Scattered Plot Graphing between GDP Growth and Taxes

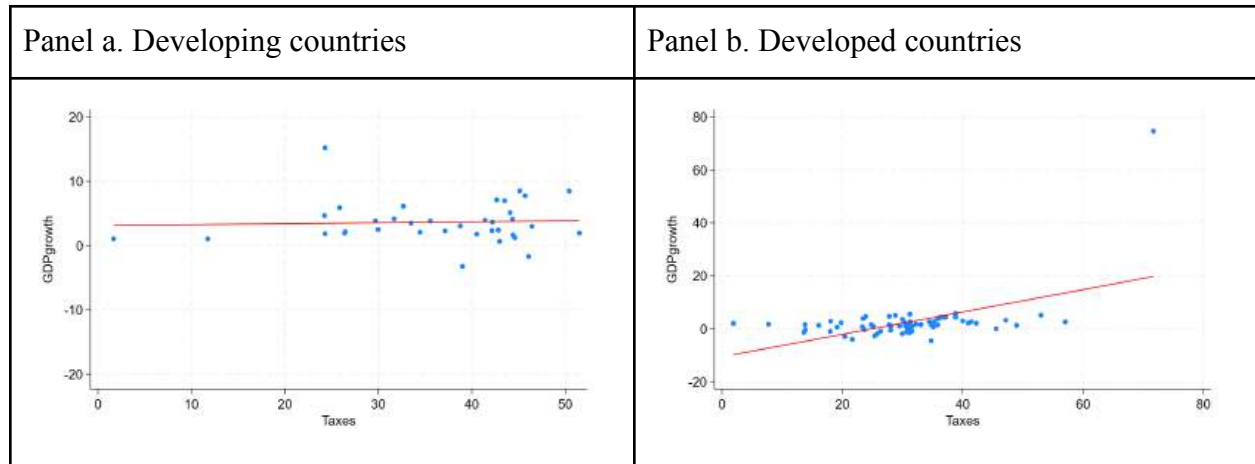


Figure 1 shows the relationship between taxes and GDP growth for developing countries in panel a and developed countries in panel b. In developing countries (panel a), the points are widely scattered with no clear pattern, suggesting that taxes have little or no impact on GDP growth. In contrast, in developed countries (panel b) there is a slightly upward trend indicating higher taxes are associated with GDP growth. This might be because taxes fund infrastructure, education, and welfare which boost productivity and growth. As the economy gets bigger, it naturally leads to higher tax revenue. This is supported by regression results in developed countries.

Figure 2. Scattered Plot Graphing between Inflation and Taxes

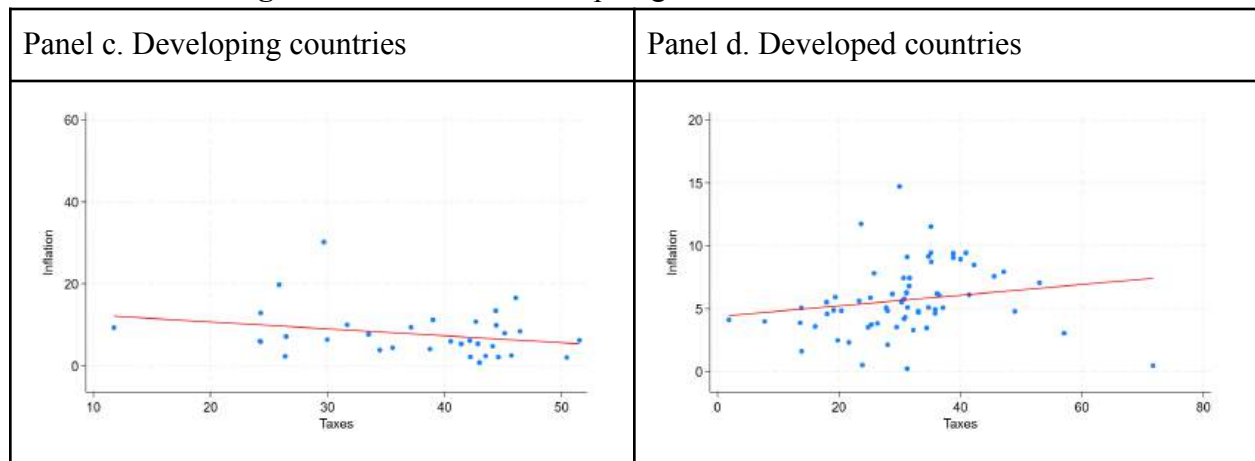


Figure 2 shows the relationship between taxes and inflation for developing countries in panel c and developed countries in panel d. In developing countries (panel c), taxes have a slightly negative effect on inflation. This might be because when taxes increase, people have less money to spend, lowering demand and when the demand drops, prices go down which reduces inflation. In contrast, for developed countries, taxes have a slightly positive effect on inflation. This might be because if business taxes go up, companies have to spend more, they transfer this cost to consumers by increasing prices, which leads to higher rates of inflation.

Figure 3. Scattered Plot Graphing between FDI Inflows and Taxes

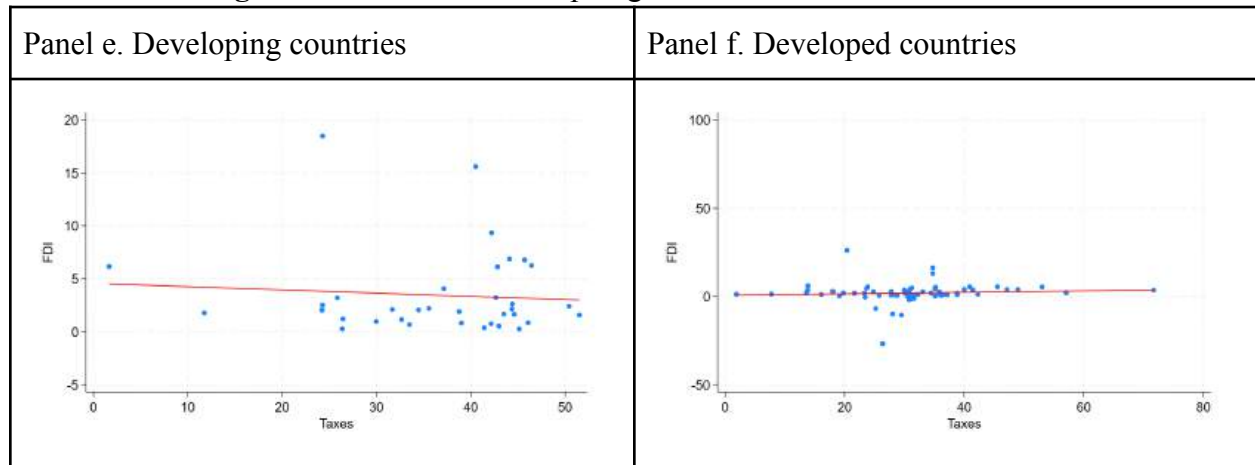


Figure 3 shows the relationship between taxes and FDI inflows for developing countries in panel e and developed countries in panel f. Taxes don't have a strong effect on attracting foreign direct investment for both developed (panel e) and developing countries (panel f). This might be because investors look at the bigger picture, not just taxes, but factors like political stability, infrastructure, and market size. Basically, a good business environment attracts international funds more than low taxes do.

Figure 4. Scattered Plot Graphing between Labor Force and Taxes

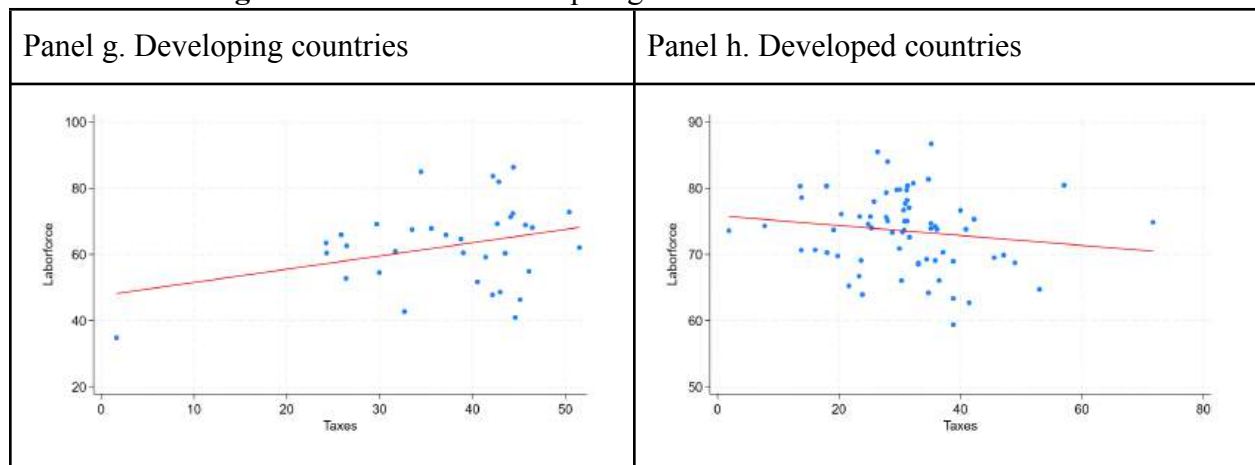


Figure 4 shows the relationship between taxes and labor force for developing countries in panel g and developed countries in panel h. Taxes have a strong positive effect on labor force for developing countries (panel h) while having a strong negative effect on developed countries (panel g). In developing countries, incomes are generally lower if taxes go up, people might feel the need to work more hours to pay bills and survive. In developed countries, people are generally wealthier, so they have more choices, so people might choose to work fewer hours, retire earlier, or rely on welfare because working extra isn't worth the tax cost.

Figure 5. Scattered Plot Graphing between Life Expectancy and Taxes

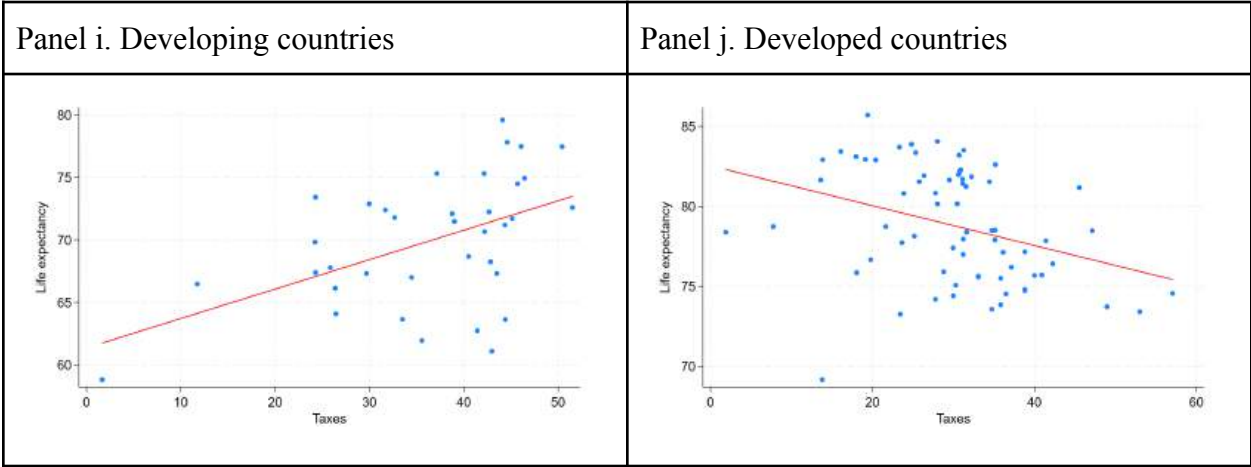
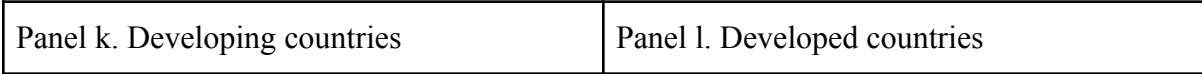


Figure 5 shows the relationship between taxes and life expectancy for developing countries in panel i and developed countries in panel j. Taxes have a strong positive effect on life expectancy for developing countries (panel i) while having a strong negative effect on developed countries (panel j). In developing countries, taxes are helpful because the governments use the money for hospitals and providing clean water, etc since countries need better healthcare, spending tax revenue on it directly increases life expectancy. In developed countries, if taxes get too high, people have less money to spend on themselves, this means they might not be able to afford things that keep them healthy which can lower life expectancy.

Figure 6. Scattered Plot Graphing between Unemployment and Taxes



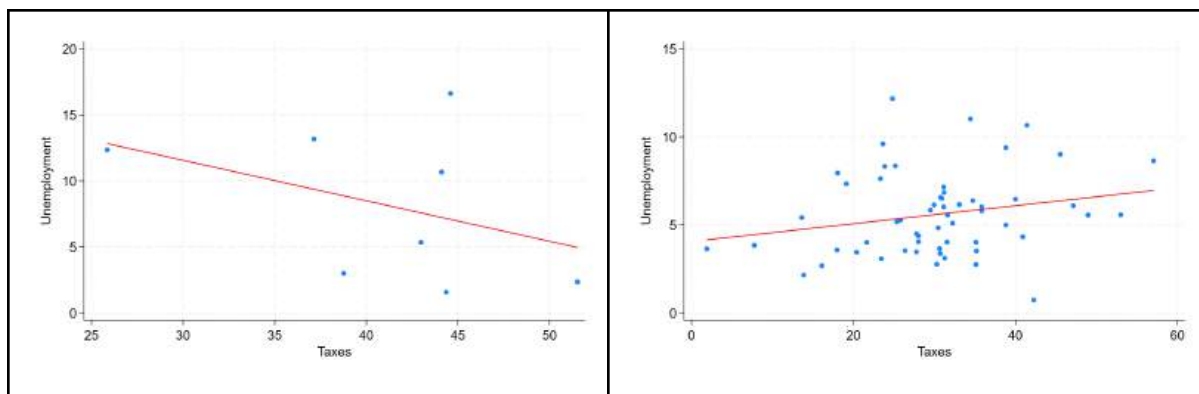


Figure 6. Scattered Plot Graphing between Unemployment and Taxes Figure 6 shows the relationship between taxes and unemployment for developing countries in panel k and developed countries in panel l. In developing countries (panel k), taxes have a strong negative effect on unemployment. In contrast, in developed countries (panel l) taxes have a slightly positive effect on unemployment. For developing countries, the government uses tax revenue to start big projects like building roads which helps in creating more jobs, lowering unemployment. For developed countries, higher taxes can make it expensive for businesses to hire people if the cost is too much to employ someone, companies hire fewer workers, which supports unemployment.

Figure 7. Scattered Plot Graphing between Current Account Balance and Taxes

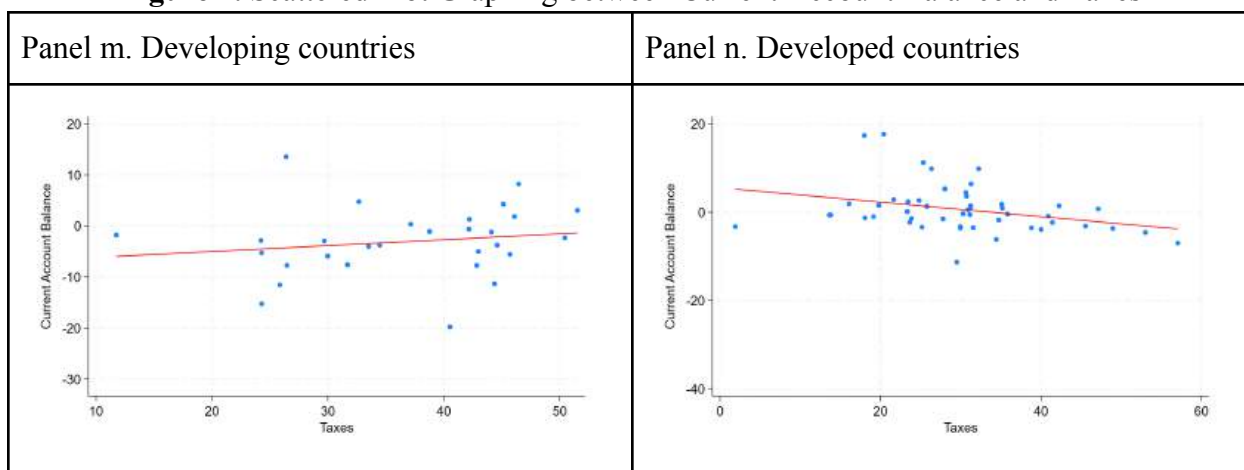


Figure 7 shows the relationship between taxes and current account balance for developing countries in panel m and developed countries in panel n. In developing countries (panel m), taxes have a slightly positive effect on current account balance. In contrast, in developed countries (panel n) taxes have a strong negative effect on current account balance. In developing countries when the taxes go up, people have less extra money to spend, this means they buy fewer imported goods since the country is spending less on imports, its current account balance grows. In developed countries, higher taxes are an extra cost for businesses, this makes the product more expensive if their goods become expensive, the difficulty to sell to other countries increase,

which means exports go down when the exports go down, so the current account balance decreases.

Figure 8. Scattered Plot Graphing between Government Effectiveness and Taxes

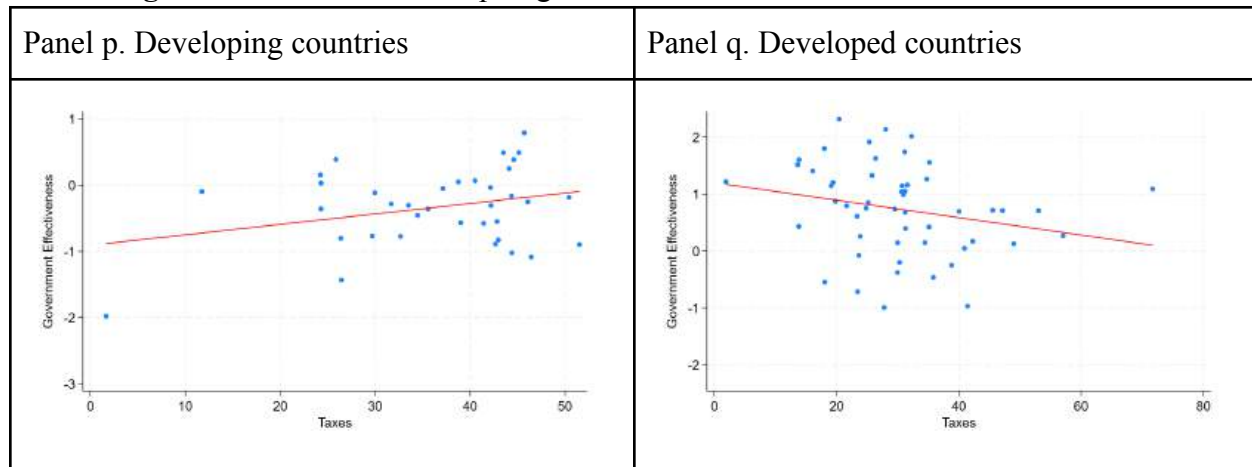


Figure 8 shows the relationship between taxes and government effectiveness for developing countries in panel p and developed countries in panel q. In developing countries (panel p), taxes have a strong positive effect on government effectiveness. In contrast, in developed countries (panel q) taxes have a strong negative effect on government effectiveness. As the government's fund comes from taxes when there is a greater amount of tax, governments are able to provide better services which fosters more trust and accountable government. For developed countries, higher taxes can slow a country's economy which reduces the amount of money the government can spend, so if the spent money doesn't satisfy the peoples' expectations, government effectiveness will worsen or in other words, their ability to function well.

Figure 9. Scattered Plot Graphing between Political Stability and Taxes

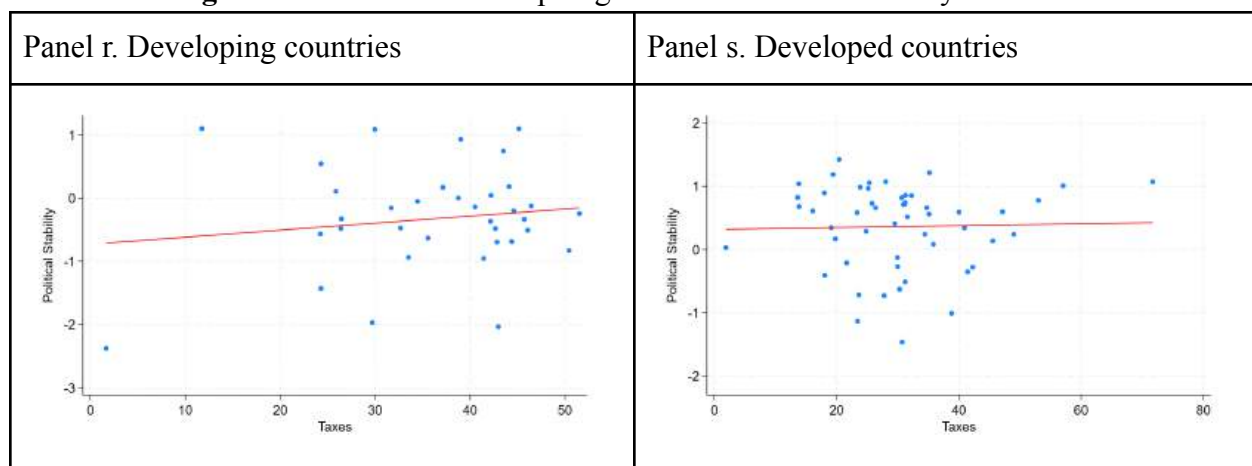


Figure 9 shows the relationship between taxes and political stability for developing countries in panel r and developed countries in panel s. In developing countries (panel r), taxes have a slightly positive effect on political stability. In contrast, in developed countries (panel s) taxes have little effect on political stability. For developing countries, taxes fund public facilities and services, which allows the government to provide services efficiently through this the government strengthens their control. For developed countries, public services and facilities have already been established, so taxes have less of an effect on political stability.

Figure 10. Scattered Plot Graphing between Rule of Law and Taxes

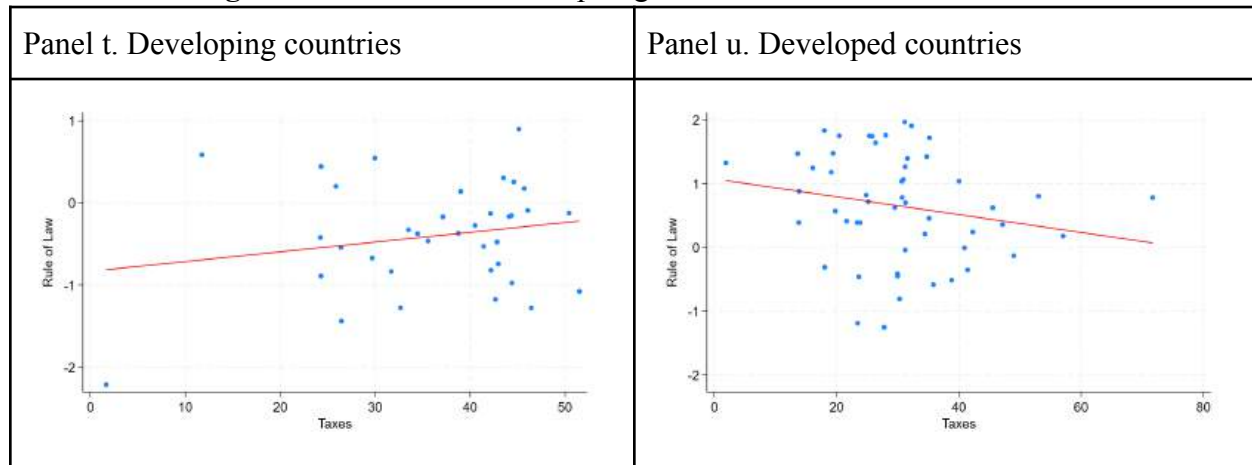


Figure 10 shows the relationship between taxes and rule of law for developing countries in panel t and developed countries in panel u. In developing countries (panel t), taxes have a strong positive effect on rule of law. In contrast, in developed countries (panel u) taxes have a strong negative effect on rule of law. For developing countries, a strong system of law ensures fairness, which increases the willingness of citizens to pay their taxes. In developed countries, higher taxes can be linked to a lower rule of law because very strong institutions allow governments to collect enough revenue without raising taxes.

Table 3a. Regression Analysis for Developed Countries

VARIABLES	(Model 1) GDP Growth	(Model 2) Inflation	(Model 3) FDI Inflows	(Model 4) Labor Force	(Model 5) Political Stability
Taxes	0.065** (0.030)	0.038 (0.047)	0.046 (0.060)	-0.056 (0.084)	0.001 (0.007)
Constant	-0.968 (0.949)	4.298*** (1.295)	0.381 (2.363)	76.291*** (2.348)	0.385 (0.232)
Observations	51	50	49	51	41
R-squared	0.081	0.025	0.005	0.010	0.000

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3a presents the results in 5 different regression models in developed countries. The independent variables are GDP, inflation, FDI, labor force, and political stability. Each column corresponds to a distinct model with explanatory variables as taxes on goods and services. If the results are statistically significant, at 1 percentage level, we indicated by (***), 5 percentage level (**), 10 percentage level (*).

In model 1, tax has positive and significant effect on GDP, with a coefficient of 0.065 ($p < 0.05$), suggesting that higher level of taxes are associated with higher levels of GDP and this positive relationship may be because higher taxes allow governments in developed countries invest more in education, infrastructure and innovation, which supports economic growth. In contrast, the effect of taxes on inflation, FDI, labor force, and political stability are statistically insignificant.

Table 3b. Regression Analysis for Developed Countries

	(Model 6)	(Model 7)	(Model 8)	(Model 9)	(Model 10)
VARIABLES	Life Expectancy	Unemployment	Current Account Balance	Government Effectiveness	Rule of Law
Taxes	-0.114** (0.044)	0.076*** (0.023)	-0.197** (0.094)	-0.022*** (0.006)	-0.017** (0.007)
Constant	83.177*** (1.397)	3.296*** (0.737)	6.866** (3.351)	1.559*** (0.222)	1.340*** (0.232)
Observations	51	46	40	41	41
R-squared	0.134	0.117	0.115	0.103	0.054

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3b presents the results in 5 different regression models in developed countries. The independent variables are life expectancy, unemployment, current account balance, government effectiveness, and rule of law. Each column corresponds to a distinct model with an explanatory model as taxes on goods and services.

Model 6 show that taxes has negative and statistically insignificant effects on life expectancy with the coefficient of -0.144 ($p < 0.05$) and on current account balance with the coefficient of -0.197 ($p < 0.05$) in model 8, indicating that higher taxes are associated with lower life expectancy and current account balance, higher taxes may reduce household income then maybe which limit access to better healthcare and living conditions, indirectly lowering life

expectancy. Higher taxes can also reduce companies' competitiveness by increasing production costs, which may lower exports and worsen the current account balance.

Taxes have a positive and significant effect on unemployment in model 7 with the coefficient of 0.076 ($p < 0.01$), suggesting that higher taxes are linked to higher unemployment levels. Higher taxes can increase labor costs for companies discouraging hiring labor demand which can raise unemployment. This may also reduce incentive for businesses to expand. In addition, taxes show a negative and significant effect on the rule of law and government effectiveness with the coefficient of -0.017 ($p < 0.01$) and -0.022 ($p < 0.01$) respectively. Very high tax levels can sometimes be associated with inefficient or overly bureaucratic governments. High taxes without visible improvements in services can lead to tax evasion, corruption and weakening of the rule of law and government effectiveness.

Table 4a. Regression Analysis for Developing Countries

VARIABLES	(Model 11) GDP Growth	(Model 12) Inflation	(Model 13) FDI Inflows	(Model 14) Labor Force	(Model 15) Political Stability
Taxes	0.021 (0.040)	-0.089 (0.072)	-0.020 (0.044)	0.184 (0.218)	0.007 (0.017)
Constant	2.663* (1.526)	10.538*** (2.909)	3.869** (1.733)	58.368*** (8.151)	-0.517 (0.653)
Observations	53	51	52	51	45
R-squared	0.006	0.027	0.004	0.026	0.009

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4a presents the results in 5 different regression models in developing countries. The independent variables are GDP, inflation, FDI, labor force, and political stability. Each column corresponds to a distinct model with an explanatory model as taxes on goods and services.

The results show that taxes have no statistically significant effect on GDP, inflation, FDI, labor force, and political stability as all coefficients are small and insignificant. These findings suggest taxes do not appear to play a substantial role in shaping macroeconomic performances in developing countries.

Table 4b. Regression Analysis for Developing Countries

VARIABLES	(Model 16) Life Expectancy	(Model 17) Unemployment	(Model 18) Current Account Balance	(Model 19) Government Effectiveness	(Model 20) Rule of Law
Taxes	0.165**	-0.267*	0.060	0.011	0.009

	(0.068)	(0.147)	(0.085)	(0.011)	(0.013)
Constant	65.739***	17.299***	-4.799	-0.723*	-0.686
	(2.536)	(5.654)	(3.244)	(0.425)	(0.522)
Observations	53	21	38	45	45
R-squared	0.113	0.231	0.011	0.050	0.024

Robust standard errors in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4b presents the results in 5 different regression models in developed countries. The independent variables are life expectancy, unemployment, current account balance, government effectiveness, and rule of law. Each column corresponds to a distinct model with an explanatory model as taxes on goods and services.

Model 16 shows that taxes have a positive and significant effect on life expectancy in developing countries with the coefficient of 0.165 ($p < 0.05$), suggesting that higher tax levels are associated with longer life expectancies. This positive relationship may be because higher tax values allow governments in developing countries to invest in more healthcare, education and social welfare programs, which can improve living conditions, increasing life expectancy. In contrast, taxes are linked to lower unemployment levels ($p < 0.10$). This negative relationship may be because higher tax values enable governments in developing countries to create jobs which can reduce unemployment.

Conclusion

The study investigates the effect of taxation of socioeconomic indicators on developing and developed countries. In developed economies, there is a positive relationship between taxation and unemployment while there is a negative relationship between taxation and government effectiveness. This suggests that in developed economies, the main challenge is getting access to a high degree of monetary resources and ensuring the government will operate at a consistent and effective level. In developing economies, there is a positive relationship between taxation and life expectancy while there is a negative relationship between taxation and unemployment. Taxation appears to have an indirect effect on social indicators through government expenditures.

It would be useful in the future to separate the data into different types of tax such as income tax, corporate tax, and sales tax. We compare developing and developed countries, but these are still very large groups, future studies might focus on specific regions like Asia, Europe, and North America. Every region has its own historical culture which definitely changes how tax affects socioeconomic indicators.

In this paper, we use regression analysis. Future research should use advanced statistical methods such as panel data analysis, GMM, and ARDL.

Works Cited

- Alfö, M., Carbonari, L., & Trovato, G. (2023). On the effects of taxation on growth: an empirical assessment. *Macroeconomic Dynamics*, 27(5), 1289-1318.
- Alshubiri, F. (2024). Do foreign direct investment inflows affect tax revenue in developed and developing countries?. *Asian Review of Accounting*, 32(5), 781-810.
- Atan, J. A. (2013). Tax policy, inflation and unemployment in Nigeria (1970–2008). *European Journal of Business and Management*, 5(15), 114-129.
- Bah, M. (2024). Tax revenue mobilization and institutional quality in sub-Saharan Africa: An empirical investigation. *African Development Review*, 36(2), 201-221.
- Bani-Mustafa, A., Nimer, K., Uyar, A., & Schneider, F. (2024). Effect of government efficiency on tax evasion: the mediating role of ethics and control of corruption. *International Journal of Public Administration*, 47(2), 90-105.
- Bashir, M. F., Sharif, A., Staniewski, M. W., Ma, B., & Zhao, W. (2024). Environmental taxes, energy transition and sustainable environmental technologies: A comparative OECD region climate change analysis. *Journal of environmental management*, 370, 122304.
- Bozatli, O., & Akca, H. (2024). Effectiveness of environmental protection expenditures and resource tax policy in the Netherland's load capacity factor: Do government effectiveness and renewable energy matter? Evidence from Fourier augmented ARDL. *Resources Policy*, 92, 105030.
- Chettri, K. K., Bhattarai, J. K., & Gautam, R. (2023). Determinants of tax revenue in South Asian countries. *Global Business Review*, 09721509231177784.
- Freund, C., & Gagnon, J. (2017). Effects of consumption taxes on real exchange rates and trade balances. *Peterson Institute for International Economics Working Paper*, (17-5).
- Karras, G., & Furceri, D. (2009). Taxes and growth in Europe. *South-Eastern Europe Journal of Economics*, 7(2).
- Malik, A., & Meraj, M. (2024). Impact of Taxation and Political Stability on Economic Growth of Pakistan. *Pakistan Journal of Humanities and Social Sciences*, 12(3), 2591-2604.
- Sánchez-Ballesta, J. P., & Yagüe, J. (2024). Tax avoidance and debt maturity in SMEs. *Journal of International Financial Management & Accounting*, 35(2), 429-464.
- Shaqiri, V., Elshani, A., & Ahmeti, S. (2024). The Effect of Direct and Indirect Taxes on Economic Growth in Developed Countries. *Ekonomika*, 103(2), 123-139.
- Shabbir, A., Abbas, S., Javed, I., & Nisar, N. (2025). The Role of Tax Policies, Inflation, Trade Openness and Exchange Rate Volatility on the Foreign Investment Decisions in Pakistan. *Indus Journal of Social Sciences*, 3(2), 636-648.
- Soto, G., Jardon, C. M., & Martinez-Cobas, X. (2024). FDI and income inequality in tax-haven countries: The relevance of tax pressure. *Economic Systems*, 48(1), 101172.
- Oluchi, N. L., & Abel, U. O. (2024). Impact of taxation on foreign direct investment in Nigeria (2000–2020). *International Journal of Innovative Finance and Economics Research*, 12(1), 143-155.

Oyeleke, O. J., & Onatunji, O. G. (2024). Macroeconomic Effects of Fiscal Policy Shocks in Nigeria: A New Proof from Mixed-Frequency Vector Autoregression Approach. *Global Business Review*, 09721509241295837.

Wang, L. (2024). Political cycle and effective corporate tax rate: evidence from China. *Applied Economics*, 56(9), 1035-1048.

Zirgulis, A., & Šarapovas, T. (2017). Impact of corporate taxation on unemployment. *Journal of Business Economics and Management*, 18(3), 412-426

Appendings

Table 1. Data definition

Indicator	Variable	Definition	Source
Taxes	Taxes on goods and services (% of revenue)	General taxes on goods and services are taxes levied on the production, leasing, delivery, sale, purchase or other change of ownership of a wide range of goods and the provision of a wide range of services.	World Bank WDI
GDP Growth	GDP per capita (annual % growth)	Gross domestic product is the total income earned through the production of goods and services in an economic territory during an accounting period.	World Bank WDI
Inflation	Inflation, consumer prices (annual % growth)	Inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly.	World Bank WDI
FDI Inflows	Foreign direct investment, net inflows (% of GDP)	Foreign direct investment are the net inflows of investment to acquire a lasting management interest (10 percent or more of voting stock) in an enterprise operating in an economy other than that of the investor.	World Bank WDI
Labor Force	Labor force participation rate, total (% of total population ages 15-64) (modeled ILO estimate)	Labor force participation rate is the proportion of the population ages 15-64 that is economically active: all people who supply labor for the production of goods and services during a specified period.	World Bank WDI
Life Expectancy	Life expectancy at birth, total (years)	Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.	World Bank WDI
Unemployment	Unemployment, total (% of total labor force) (national estimate)	Unemployment refers to the share of the labor force that is without work but available for and seeking employment.	World Bank WDI
Current Account Balance	Current account balance (% of GDP)	Balance of current transactions (transactions in goods and services, earned income and transfer income) between residents and non-residents. The term current account balance is used in the external accounts and is expressed from the perspective of resident units.	World Bank WDI
Government Effectiveness	Government Effectiveness: Estimate	Government Effectiveness captures perceptions and views of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.	World Bank WDI
Political Stability	Political Stability and Absence of Violence/Terrorism: Estimate	Political Stability and Absence of Violence/Terrorism measures perceptions of the likelihood of political instability and/or politically-motivated violence, including terrorism.	World Bank WDI
Rule of Law	Rule of Law: Estimate	Rule of Law captures perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence.	World Bank WDI
CPIA debt rating	CPIA debt policy rating (1=low to 6=high)	The Debt Policy and Management criterion assesses whether the country's debt management strategy is conducive to ensure medium-term debt sustainability and minimize budgetary risks.	World Bank WDI

Pancreatic Ductal Adenocarcinoma: Current Challenges and Emerging Therapeutic Frontiers

By Bowen Ke, Mentor By Ricardo Harripaul

Abstract

Pancreatic ductal adenocarcinoma (PDAC) remains one of the most aggressive and deadliest cancers in the world today. Its global incidence continues to rise, and survival rates remain low. Late diagnosis, high tumor heterogeneity, and a highly immunosuppressive tumor microenvironment are the main reasons for limited treatment success. This review aims to provide a comprehensive overview of PDAC, from its fundamental information including epidemiology, genetic landscape, tumor microenvironment, and current standard treatments, to emerging therapeutic strategies aimed at overcoming key biological barriers and improving clinical outcomes, such as immunotherapy, nanotechnology-based drug delivery, gene editing, and cancer vaccines.

Despite encouraging progress, numerous challenges remain, including difficulties in drug delivery due to the dense stromal barrier and individual variability in patient response. Current research focuses on combination therapies and innovative delivery systems to improve treatment outcomes. Future research should focus on precision medicine, improving clinical trial design, and advancing scalable and safe technologies for clinical application. In summary, these efforts offer hope for transforming PDAC from a highly fatal disease into one that is more manageable and potentially curable.

Introduction

Cancer continues to be a significant global health burden, with high morbidity and mortality. Among the many types of cancer, pancreatic cancer is one of the deadliest cancers worldwide and also one of the cancers with the lowest survival rates. The prognosis for pancreatic cancer remains poorer compared to other malignant tumors, with only a small number of patients surviving for more than one year (McGuire, 2016).

Pancreatic ductal adenocarcinoma (PDAC), which accounts for over 90% of pancreatic cancer cases, has a late diagnosis, and limited therapeutic options (Kleeff et al., 2016). PDAC mortality has more than doubled globally over the past 25 years—from about 196,000 cases in 1990 to approximately 441,000 in 2017 (Klein, 2021a). Incidence is high in North America, Europe, and Australia, with age-adjusted rates rising driven by aging populations and the rising prevalence of risk factors (Klein, 2021b). Specifically in the U.S., Black individuals have a higher incidence than White, Hispanic, and Asian populations—partly due to differences in smoking, obesity, and diabetes rates, which are considered significant risk factors of PDAC (Klein, 2021b).

Risk factors for pancreatic cancer can be categorized into modifiable and genetic determinants. Modifiable factors include smoking, which remains the single most significant modifiable risk worldwide. Obesity and metabolic factors also play an important role, especially type 2 diabetes, substantially increasing PDAC risk (Behrens et al., 2015). Insulin resistance,

chronic low-level inflammation, and high blood glucose levels contribute to genetic instability and cell proliferation. In addition, high levels of alcohol consumption are linked to a higher risk, likely due to its role in chronic pancreatitis and inflammation (Tramacere et al., 2010). In addition to modifiable factors, inherited genetic variants account for approximately 22–33% of PDAC risk. Hereditary cancer genes (e.g., BRCA1/2, PALB2), and genes that are associated with hereditary pancreatitis are common drivers of pancreatic cancer (Rizzato et al., 2013).

This paper's objective is to provide a comprehensive overview of pancreatic ductal adenocarcinoma (PDAC). The research paper examines the disease's epidemiology, genetic mutations, and tumor microenvironment, emphasizing how these factors contribute to poor prognosis and treatment resistance. Then it discusses current standard treatments, such as surgery and chemotherapy, and highlights emerging therapeutic strategies including immunotherapy, nanotechnology-based drug delivery, gene editing, and cancer vaccines. These innovative approaches aim to improve drug delivery, immune activation, and precision targeting.

Genetic Landscape

The genetic basis of pancreatic ductal adenocarcinoma (PDAC) is determined by several mutations that drive tumor development. The most common mutations occur in the KRAS oncogene, affecting approximately 90% of cases (Jones et al., 2008). Tumor suppressor genes are also frequently inactivated: CDKN2A (~98%) disrupts cell cycle regulation; TP53 (~70%) impairs DNA repair and apoptosis; and SMAD4 (~50%) alters TGF- β signaling, thereby promoting tumor invasion (Australian Pancreatic Cancer Genome Initiative et al., 2015). These mutations collectively contribute to the aggressive clinical presentation and poor prognosis of PDAC.

These mutations influence therapeutic response and highlight opportunities for precision medicine, which tailors medical treatments to individual patients based on their unique genetic makeup. For patients with homologous recombination deficiency (HRD), mostly caused by BRCA1/2 or PALB2 mutations, predicts increased sensitivity to platinum-based chemotherapy or PARP inhibitors (Park et al., 2020). Therefore, PARP inhibitors were approved as therapies for this patient population.

For patients with KRAS mutations (e.g., G12D/G12V), downstream pathway components such as SHP2 and SOS1 are emerging targets with promising inhibitors in development (Brozos-Vázquez et al., 2024). In addition, other genomic alterations such as gene fusion or translocation present potentially actionable targets that allow patients to achieve positive treatment responses. A gene fusion occurs when two previously separated genes combine abnormally, creating a new hybrid gene that can drive cancer development. This hybrid gene can produce abnormal proteins, and the resulting fusion protein may exhibit uncontrolled activity, such as signaling the cell to grow and divide continuously. Gene fusions can also place growth-promoting genes under the control of highly activated promoters, leading to overexpression and loss of normal cell regulation, which can contribute to tumor development (Dai et al., 2018). In PDAC, although gene fusions are rare, accounting for less than 1% of

patients, patients with fusions such as NTRK, ALK, and NRG1 may respond dramatically to targeted drugs that already exist. Therefore, identifying gene fusions could open some powerful treatment paths for this subset of PDAC patients (Nevala-Plagemann et al., 2020).

Tumor Microenvironment

Understanding and overcoming barriers in the tumor microenvironment (TME) can significantly enhance patients' clinical treatment response. TME signatures include desmoplasia level, T-cell infiltration, and high immune-suppressive cell ratios; together, these features pose significant challenges to current treatment strategies (Farhangnia et al., 2024). A primary component of the TME is stromal tissue, which is itself composed of cancer-associated fibroblasts (CAFs), tumor-associated macrophages (TAMs), myeloid-derived suppressor cells (MDSCs), dendritic cells (DCs), natural killer (NK) cells, and other innate immune cells (Farhangnia et al., 2024). When stromal cells replace normal pancreatic tissue, it results in fibrotic proliferation, a defining feature of PDAC. Fibrotic proliferation creates a physical barrier that limits the penetration and efficacy of chemotherapy and immunotherapy (Jacobetz et al., 2013). The TME also restricts T-cell infiltration and activation in order to suppress immune responses against the tumor. The TME immune landscape in PDAC is shaped by abundant immunosuppressive cells. The low tumor mutational burden and poor antigenicity reduce the likelihood of eliciting strong immune recognition (Farhangnia et al., 2024). Therefore, it is difficult for the immune system to mount an effective immune response against the tumor.

Biomarkers

Biomarkers play an essential role in improving the diagnosis, prognosis, and treatment of PDAC. As mentioned earlier, inactivating mutations in HR genes suggest heightened sensitivity to PARP inhibitor treatment. Tumor tissue-based transcriptomic and proteomic analyses have identified distinct gene-expression subtypes, including classical and basal-like forms, which lead to differences in patients' chemosensitivity. Traditional circulating protein markers like carbohydrate antigen 19-9 (CA19-9), carcinoembryonic antigen (CEA), and cancer antigen 125 (CA125) are used to monitor disease, but their lack of specificity limits their effectiveness in early detection (Brozos-Vázquez et al., 2024). Recent advances in circulating biomarkers offer new and non-invasive opportunities. For instance, in the blood, circulating tumor DNA (ctDNA) and RNA (ctRNA) can be used to detect genomic alterations, such as mutations, rearrangements or fusions. Additionally, exosomal proteins and microRNAs (miRNAs) can reflect tumor activity, and have early-detection potential. Circulating tumor cells (CTCs), although lower in abundance, may also help monitor metastasis risk (Brozos-Vázquez et al., 2024).

Beyond tumor-intrinsic signals, biomarkers derived from the TME are increasingly important. TME signatures, including PD-L1 expression, desmoplasia level, T-cell infiltration, and immune-suppressive cell ratios, can help predict immunotherapy responses (Whatcott et al., 2015). However, limitations remain. Many biomarkers, particularly circulating ones, need better validation in large cohorts. In addition, tumor heterogeneity precludes a single biomarker from

capturing the full complexity of PDAC (Brozos-Vázquez et al., 2024). Future diagnostic strategies will increasingly depend on multi-omics biomarker development, incorporating genomic, transcriptomic, proteomic, and immune profiling to identify clinically actionable alterations and stratify patients more precisely. This approach heightens the need for scalable, affordable assays for clinical deployment, with the potential to transform PDAC diagnosis from a uniform model to a precision biomarker-driven framework that enables personalized treatment selection.

Current Standard-of-Care Treatments

The current standard treatments for PDAC still have limitations. Surgery and chemotherapy are the two most common treatments for PDAC. Surgery is currently the only curative treatment, yet less than 20% of patients are resectable due to late detection and rapid disease progression (He et al., 2014). Chemotherapy is the standard treatment for most patients but is not curative, serving instead to slow disease progression and extend survival. Even after resection, chemotherapy is commonly used to reduce the risk of recurrence. Drugs, including FOLFIRINOX and the combinations of gemcitabine with nab-paclitaxel, have demonstrated improvements in patients' survival (Conroy et al., 2011). Additionally, targeted therapy is beginning to enter clinical practice. As discussed previously, PARP inhibitors have shown effectiveness for BRCA-mutated PDAC patients, representing an essential step in biomarker-guided treatments (Kleeff et al., 2016). Although remarkable progress has been made, therapy outcomes remain poor, and patients need more effective and personalized strategies.

Immunotherapy

Immunotherapy has improved treatments for many types of cancers; however, its application to PDAC has had limited success when used as a monotherapy. There are many types of immunotherapy approaches. Immune checkpoint inhibitors (ICIs), which are drugs that block immune checkpoints, targeted PD-1/PD-L1 or CTLA-4, work in only a small number of patients with special tumor features (Ye et al., 2024). This means that less than 2% of PDAC patients show meaningful responses. However, combinations of other therapies with immune checkpoint inhibitors, – for example, PARP plus PD-1 inhibitors – have the potential to be more effective (Kleeff et al., 2016). Such combinations could enhance immune recognition of tumor cells. Similarly, oncolytic virus therapies, which selectively lyse tumor cells and stimulate immune responses, could be seen as a means to overcome the immune resistance of PDAC (Farhangnia et al., 2024).

Adoptive cell therapies are another promising path. Genetically engineered T cells, such as the TCR-engineered T cells, can target specific tumor antigens. In addition to the chimeric antigen receptor (CAR) T cells, which specifically target pancreatic tumor antigens, CAR-NK cell therapy is another strategy (Zhao et al., 2019). CAR-NK cells are genetically engineered to express CARs, which link antibodies or receptors that recognize surface antigens on target cells, such as cancer cells, to signaling molecules that activate immune cells. This modification can

counteract inhibitory receptors, therefore enhancing the killing activity of NK cells against target cells (Wang et al., 2024). At the same time, cytokine-induced killer (CIK) cells provide cytotoxicity with nonspecific antitumor effects. Although these therapies are still in experimental stages, they represent innovative pathways to reprogram immune responses against PDAC.

Novel Drug Delivery Method Using Nanotechnology

Nanotechnology has immense potential to improve treatment responses from PDAC patients. It offers innovative strategies to overcome many challenges associated with PDAC, especially the dense stroma, poor drug penetration, and multidrug resistance associated with TME. Common nanocarriers, ranging in size from 1–200 nm, include liposomes, micelles, polymeric nanoparticles, and solid lipid carriers (Luo & Zhang, 2024). They are designed to improve drug solubility, stability, and delivery. Active targeting strategies, for example, decorate nanoparticles with surface ligands, such as antibodies against CD44, enable specific delivery to tumor or cancer stem cells (CSCs). Additionally, stimuli-responsive nanocarriers can release their payload under specific conditions, such as the acidic, enzyme-rich tumor environment. This ensures drug release within the TME, enhancing drug efficacy and reducing systemic toxicity (Luo & Zhang, 2024).

One of the key advantages of nanotechnology is its ability to overcome mechanisms of treatment resistance in PDAC. Stroma targeted nanoparticles co-delivering stromal-degrading enzymes (e.g., matrix metalloproteinases, MMPs) or fibroblast-targeting agents help remodel the extracellular matrix (ECM), therefore softening the physical barrier of TME and enhancing drug penetration (Luo & Zhang, 2024). Another design focuses on targeting cancer stem cells (CSCs). For instance, olive oil-based nanocapsules decorated with anti-CD44 antibodies and loaded with paclitaxel achieved nearly 4 times higher efficacy against pancreatic CSCs in vivo (Luo & Zhang, 2024). In addition, nanocarriers can modulate the immune system. They are being developed to deliver immunomodulatory agents, such as STING or TLR agonists, which increase antigen presentation and T-cell infiltration (Luo & Zhang, 2024). All three designs not only improve drug delivery but also modulate the immunosuppressive environment of the TME.

Preclinical and early studies have provided strong, encouraging evidence of nanotechnology's potential in PDAC therapy. Animal models show robust anti-tumor efficacy and survival benefits, especially for systems combining chemotherapy with stromal remodeling or CSC targeting. Examples include magnetoliposomes loaded with cisplatin, which induce hyperthermia under electromagnetic fields and improve tumor regression and survival in mice (Luo & Zhang, 2024).

Despite numerous preclinical successes being strong, few nanocarrier platforms have progressed to clinical trials specific to PDAC. This indicates the gap between laboratory experiments and clinical translation. Therefore, continued optimization, large-scale clinical trials, and cost-effective manufacturing are crucial to integrating nanotechnology into the treatment of PDAC.

Gene Editing

The field of gene editing has evolved rapidly over the past two decades, with each subsequent generation of tools improving precision and usability. Gene editing introduces precise modifications into the genome via insertions, deletions, or base substitutions (Li et al., 2023). Early forms of therapy include zinc-finger nucleases (ZFNs), and transcription activator-like effector nucleases (TALENs). They are protein-based targeted cuts that rely on engineered proteins to break the targeted DNA double-strand; however, they require complex designs and customization. CRISPR/Cas9 is the third generation of gene-editing tools and has been widely adopted in recent years. It revolutionized gene editing by enabling modular design of gene editing complexes, such that the only component that needs to be customized is the guide RNA for specific targeting, which can be easily exchanged (Hsu et al., 2014). The Cas 9 enzyme is directed by the single-guide RNA (sgRNA) to a specific DNA sequence adjacent to a protospacer-adjacent motif (PAM), where it creates a DNA double-strand break (Li et al., 2023). This new generation of gene editing tools is simpler, cheaper, and more adaptable than the first two generations.

Additionally, even more advanced versions are now being developed. For example, base editors can convert one DNA base to another, enabling precise single-base changes without double-strand breaks. Prime editors can introduce targeted insertions or corrections of longer sequences by using reverse transcriptase and a special template. In addition to greater precision, it has fewer safety concerns (Li et al., 2023).

However, although these tools work well in vitro, safe and efficient delivery of gene-editing tools into an organism's cells is one of the biggest hurdles in the delivery system. Viral vectors, especially adeno-associated viruses, can deliver editing components efficiently but also pose risks of immune reactions and have size limits (Li et al., 2023). In comparison, non-viral systems, including lipid nanoparticles and polymeric carriers, offer a promising alternative. They can be engineered to recognize target cells specifically, therefore increasing specificity and safety. One key focus for future research should be the development of disease-targeted non-viral delivery systems, possibly using the unique physiological changes in diseased tissue as targeting cues. Clinical research has shown the feasibility of this approach. The first clinical trial for CRISPR in 2016 successfully edited T cells to reduce PD-1 expression (Lu et al., 2020). Beyond immuno-oncology, CRISPR and related technologies have potential in applications that include treatment of genetic disorders, cardiovascular disease, and neurodegeneration. However, other challenges remain. For example, off-target effects are unintended edits that can cause genomic instability. Large deletions or translocations from misrepair could cause chromosomal issues. In addition, some regions are inaccessible due to strict PAM requirements for Cas9.

Cancer Vaccines

Parallel to advances in gene editing technology, cancer vaccines have emerged as a complementary strategy to stimulate the immune system against tumors. Vaccines are classified

into two types: therapeutic and prophylactic. Therapeutic vaccines are administered after cancer onset and aim to eliminate existing tumors and build lasting immune memory. Conversely, prophylactic vaccines are administered to cancer-free individuals to generate antibodies and memory cells, thereby preventing tumor development (Ghattas et al., 2021). An example of prophylactic vaccines is FDA-approved HPV and HBV vaccines, demonstrating the potential to reduce virus-associated cancers. For the therapeutic side, dendritic cell (DC) vaccine Sipuleucel-T was the first FDA-approved cancer vaccine. Vaccine platforms under investigation include peptide vaccines, nucleic acid vaccines, viral vector vaccines, cell-based vaccines. Peptide vaccines simulate tumor antigens to stimulate T-cell immunity, with high specificity, low toxicity, and ease of production. Nucleic acid vaccines (DNA/mRNA) deliver genetic instructions for antigen proteins, have strong immunogenicity, and cost-effective production. Viral vector vaccines use viruses (e.g., adenovirus, vaccinia) to deliver antigens, often with intrinsic adjuvant properties. Cell-based vaccines typically are from patient-derived cells (e.g., dendritic cells), presented in the therapeutic category (Sipuleucel-T) (Zhou et al., 2025). Notably, mRNA vaccine BNT122 (autogene cevumeran) in PDAC combined with immunotherapy showed strong T-cell responses and delayed relapse.

Despite progress made, resistance remains a critical hurdle in cancer vaccine therapy. Tumors may evade vaccine-induced responses due to immune suppression within the microenvironment. This limits vaccines efficacy as a monotherapy. Therefore, combination strategies are getting more and more attention. Pairing vaccines with standard treatments, checkpoint inhibitors and immunotherapies (such as ICIs or ACT) can enhance efficacy and overcome resistance. Such strategies hold promise especially in solid tumors (Zhou et al., 2025). Future research should focus on optimizing antigen selection for better specificity and away from immune evasion, and refining adjuvant formulations and delivery routes for stronger immunity. Furthermore, advancing diverse vaccine platforms to address the needs of different tumor types and patients could enhance the versatility of cancer vaccines (Zhou et al., 2025). All together, these advances highlight the role of gene editing technology and cancer vaccines in reshaping the treatment landscape of PDAC and other cancer types.

Conclusion

Pancreatic cancer remains one of the most lethal malignancies, with limited treatment options and poor survival outcomes. Surgery and chemotherapy are the two main treatments now, the vast majority of patients are diagnosed at advanced stages, leading to poor treatment outcomes (Kleeff et al., 2016). In recent years, novel therapeutic strategies including nanotechnology-based drug delivery, gene editing, and cancer vaccines, have opened the path for more accurate and effective treatments. They are aimed to overcome some of the challenges that make pancreatic cancer so resistant to traditional treatments, such as the dense physical barrier and immunosuppressive environment of TME (Clark et al., 2007).

Despite the advancements of these new therapies, significant efforts need to be done before these innovations can be turned into routine clinical care. Many promising therapeutic

strategies remain in early-stage development or clinical trials. Problems associated with delivery efficiency, treatment personalization, and production scalability still require feasible solutions. By integrating breakthroughs in immunotherapy, nanomedicine, and genetic technology with smarter clinical strategies, more effective and enduring treatments are being steadily advanced. As a disease once considered nearly incurable, the hope is that pancreatic cancer will eventually become a manageable and even curable disease through increasingly advanced technology.

Works Cited

- Australian Pancreatic Cancer Genome Initiative, Waddell, N., Pajic, M., Patch, A.-M., Chang, D. K., Kassahn, K. S., Bailey, P., Johns, A. L., Miller, D., Nones, K., Quek, K., Quinn, M. C. J., Robertson, A. J., Fadlullah, M. Z. H., Bruxner, T. J. C., Christ, A. N., Harliwong, I., Idrisoglu, S., Manning, S., ... Grimmond, S. M. (2015). Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature*, 518(7540), 495–501.
<https://doi.org/10.1038/nature14169>
- Behrens, G., Jochem, C., Schmid, D., Keimling, M., Ricci, C., & Leitzmann, M. F. (2015). Physical activity and risk of pancreatic cancer: A systematic review and meta-analysis. *European Journal of Epidemiology*, 30(4), 279–298.
<https://doi.org/10.1007/s10654-015-0014-9>
- Brozos-Vázquez, E., Toledano-Fonseca, M., Costa-Fraga, N., García-Ortiz, M. V., Díaz-Lagares, Á., Rodríguez-Ariza, A., Aranda, E., & López-López, R. (2024). Pancreatic cancer biomarkers: A pathway to advance in personalized treatment selection. *Cancer Treatment Reviews*, 125, 102719. <https://doi.org/10.1016/j.ctrv.2024.102719>
- Clark, C. E., Hingorani, S. R., Mick, R., Combs, C., Tuveson, D. A., & Vonderheide, R. H. (2007). Dynamics of the Immune Reaction to Pancreatic Cancer from Inception to Invasion. *Cancer Research*, 67(19), 9518–9527.
<https://doi.org/10.1158/0008-5472.CAN-07-0175>
- Conroy, T., Desseigne, F., Ychou, M., Bouché, O., Guimbaud, R., Bécouarn, Y., Adenis, A., Raoul, J.-L., Gourgou-Bourgade, S., De La Fouchardière, C., Bennouna, J., Bachet, J.-B., Khemissa-Akouz, F., Péré-Vergé, D., Delbaldo, C., Assenat, E., Chauffert, B., Michel, P., Montoto-Grillot, C., & Ducreux, M. (2011). FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer. *New England Journal of Medicine*, 364(19), 1817–1825.
<https://doi.org/10.1056/NEJMoa1011923>
- Dai, X., Theobard, R., Cheng, H., Xing, M., & Zhang, J. (2018). Fusion genes: A promising tool combating cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, 1869(2), 149–160. <https://doi.org/10.1016/j.bbcan.2017.12.003>
- Farhangnia, P., Khorramdelazad, H., Nickho, H., & Delbandi, A.-A. (2024). Current and future immunotherapeutic approaches in pancreatic cancer treatment. *Journal of Hematology & Oncology*, 17(1), 40. <https://doi.org/10.1186/s13045-024-01561-6>
- Ghattas, M., Dwivedi, G., Lavertu, M., & Alameh, M.-G. (2021). Vaccine Technologies and Platforms for Infectious Diseases: Current Progress, Challenges, and Opportunities. *Vaccines*, 9(12), 1490. <https://doi.org/10.3390/vaccines9121490>
- He, J., Ahuja, N., Makary, M. A., Cameron, J. L., Eckhauser, F. E., Choti, M. A., Hruban, R. H., Pawlik, T. M., & Wolfgang, C. L. (2014). 2564 resected periampullary adenocarcinomas at a single institution: Trends over three decades. *HPB*, 16(1), 83–90.
<https://doi.org/10.1111/hpb.12078>
- Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, 157(6), 1262–1278.

- <https://doi.org/10.1016/j.cell.2014.05.010>
- Jacobetz, M. A., Chan, D. S., Neesse, A., Bapiro, T. E., Cook, N., Frese, K. K., Feig, C., Nakagawa, T., Caldwell, M. E., Zecchini, H. I., Lolkema, M. P., Jiang, P., Kultti, A., Thompson, C. B., Maneval, D. C., Jodrell, D. I., Frost, G. I., Shepard, H. M., Skepper, J. N., & Tuveson, D. A. (2013). Hyaluronan impairs vascular function and drug delivery in a mouse model of pancreatic cancer. *Gut*, 62(1), 112–120. <https://doi.org/10.1136/gutjnl-2012-302529>
- Jones, S., Zhang, X., Parsons, D. W., Lin, J. C.-H., Leary, R. J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., Hong, S.-M., Fu, B., Lin, M.-T., Calhoun, E. S., Kamiyama, M., Walter, K., Nikolskaya, T., Nikolsky, Y., Hartigan, J., ... Kinzler, K. W. (2008). Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses. *Science*, 321(5897), 1801–1806. <https://doi.org/10.1126/science.1164368>
- Kleeff, J., Korc, M., Apte, M., La Vecchia, C., Johnson, C. D., Biankin, A. V., Neale, R. E., Tempero, M., Tuveson, D. A., Hruban, R. H., & Neoptolemos, J. P. (2016). Pancreatic cancer. *Nature Reviews Disease Primers*, 2(1), 16022. <https://doi.org/10.1038/nrdp.2016.22>
- Klein, A. P. (2021a). Pancreatic cancer epidemiology: Understanding the role of lifestyle and inherited risk factors. *Nature Reviews Gastroenterology & Hepatology*, 18(7), 493–502. <https://doi.org/10.1038/s41575-021-00457-x>
- Klein, A. P. (2021b). Pancreatic cancer epidemiology: Understanding the role of lifestyle and inherited risk factors.
- Li, T., Yang, Y., Qi, H., Cui, W., Zhang, L., Fu, X., He, X., Liu, M., Li, P., & Yu, T. (2023). CRISPR/Cas9 therapeutics: Progress and prospects. *Signal Transduction and Targeted Therapy*, 8(1), 36. <https://doi.org/10.1038/s41392-023-01309-7>
- Lu, Y., Xue, J., Deng, T., Zhou, X., Yu, K., Deng, L., Huang, M., Yi, X., Liang, M., Wang, Y., Shen, H., Tong, R., Wang, W., Li, L., Song, J., Li, J., Su, X., Ding, Z., Gong, Y., ... Mok, T. (2020). Safety and feasibility of CRISPR-edited T cells in patients with refractory non-small-cell lung cancer. *Nature Medicine*, 26(5), 732–740. <https://doi.org/10.1038/s41591-020-0840-5>
- Luo, W., & Zhang, T. (2024). The new era of pancreatic cancer treatment: Application of nanotechnology breaking through bottlenecks. *Cancer Letters*, 594, 216979. <https://doi.org/10.1016/j.canlet.2024.216979>
- McGuire, S. (2016). *World Cancer Report 2014*. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015. *Advances in Nutrition*, 7(2), 418–419. <https://doi.org/10.3945/an.116.012211>
- Nevala-Plagemann, C., Hidalgo, M., & Garrido-Laguna, I. (2020). From state-of-the-art treatments to novel therapies for advanced-stage pancreatic cancer. *Nature Reviews Clinical Oncology*, 17(2), 108–123. <https://doi.org/10.1038/s41571-019-0281-6>
- Park, W., Chen, J., Chou, J. F., Varghese, A. M., Yu, K. H., Wong, W., Capanu, M.,

- Balachandran, V., McIntyre, C. A., El Dika, I., Khalil, D. N., Harding, J. J., Ghalehsari, N., McKinnell, Z., Chalasani, S. B., Makarov, V., Selenica, P., Pei, X., Lecomte, N., ... O'Reilly, E. M. (2020). Genomic Methods Identify Homologous Recombination Deficiency in Pancreas Adenocarcinoma and Optimize Treatment Selection. *Clinical Cancer Research*, 26(13), 3239–3247. <https://doi.org/10.1158/1078-0432.CCR-20-0418>
- Rizzato, C., Campa, D., Pezzilli, R., Soucek, P., Greenhalf, W., Capurso, G., Talar-Wojnarowska, R., Heller, A., Jamrozak, K., Khaw, K.-T., Key, T. J., Bambi, F., Landi, S., Mohelnikova-Duchonova, B., Vodickova, L., Büchler, M. W., Bugert, P., Vodicka, P., Neoptolemos, J. P., ... Canzian, F. (2013). ABO blood groups and pancreatic cancer risk and survival: Results from the PANcreatic Disease ReseArch (PANDoRA) consortium. *Oncology Reports*, 29(4), 1637–1644. <https://doi.org/10.3892/or.2013.2285>
- Sherman, M. H., & Beatty, G. L. (2023). Tumor Microenvironment in Pancreatic Cancer Pathogenesis and Therapeutic Resistance. *Annual Review of Pathology: Mechanisms of Disease*, 18(1), 123–148. <https://doi.org/10.1146/annurev-pathmechdis-031621-024600>
- Tramacere, I., Scotti, L., Jenab, M., Bagnardi, V., Bellocchio, R., Rota, M., Corrao, G., Bravi, F., Boffetta, P., & La Vecchia, C. (2010). Alcohol drinking and pancreatic cancer risk: A meta-analysis of the dose-risk relation. *International Journal of Cancer*, 126(6), 1474–1486. <https://doi.org/10.1002/ijc.24936>
- Wang, W., Liu, Y., He, Z., Li, L., Liu, S., Jiang, M., Zhao, B., Deng, M., Wang, W., Mi, X., Sun, Z., & Ge, X. (2024). Breakthrough of solid tumor treatment: CAR-NK immunotherapy. *Cell Death Discovery*, 10(1), 40. <https://doi.org/10.1038/s41420-024-01815-9>
- Whatcott, C. J., Diep, C. H., Jiang, P., Watanabe, A., LoBello, J., Sima, C., Hostetter, G., Shepard, H. M., Von Hoff, D. D., & Han, H. (2015). Desmoplasia in Primary Tumors and Metastatic Lesions of Pancreatic Cancer. *Clinical Cancer Research*, 21(15), 3561–3568. <https://doi.org/10.1158/1078-0432.CCR-14-1051>
- Ye, X., Yu, Y., Zheng, X., & Ma, H. (2024). Clinical immunotherapy in pancreatic cancer. *Cancer Immunology, Immunotherapy: CII*, 73(4), 64. <https://doi.org/10.1007/s00262-024-03632-6>
- Zhao, Z., Zheng, L., Chen, W., Weng, W., Song, J., & Ji, J. (2019). Delivery strategies of cancer immunotherapy: Recent advances and future perspectives. *Journal of Hematology & Oncology*, 12(1), 126. <https://doi.org/10.1186/s13045-019-0817-3>
- Zhou, Y., Wei, Y., Tian, X., & Wei, X. (2025). Cancer vaccines: Current status and future directions. *Journal of Hematology & Oncology*, 18(1), 18. <https://doi.org/10.1186/s13045-025-01670-w>

Digital Inequality in Schools: The Impact of School-Level Digital Accessibility on Student Participation and Psychological Isolation in Kenya and Cambodia

By Minji Choi

Abstract

This study investigates the relationship between digital accessibility at the school level and student participation and psychological isolation in elementary education in low-resource settings. The study conducted in Kenya and Cambodia used a mixed-methods approach to examine digital accessibility and its relationship to student participation and psychological isolation. Quantitative and qualitative data were collected from student and teacher surveys and logged data from online tutoring sessions. The three dimensions of digital accessibility—devices, internet, and technical failures—were used to develop a comprehensive measurement framework. Students and teachers were unable to access digital content and services that could have supported their learning due to a number of reasons, mainly to do with the stability and consistency of access. Many students reported that they experienced technical failures on a regular basis that interfered with their learning. Teachers reported that technical failure was a major cause of learning and teaching difficulties. In contrast, students in relatively stable learning environments reported that they engaged more in lessons, attended more regularly, and submitted more work on time. Unstable digital conditions caused frustration, decreased confidence, and led to students feeling disconnected from the learning space. Approaching the digital divide from a school-oriented perspective, the paper positions the issue within the institutional sphere. Addressing the digital divide at the school level improves learning and well-being by boosting engagement of students from low-income backgrounds through improved digital infrastructure.

1. Introduction

Achieving reliable internet connectivity can still be a major challenge in many low-resource areas, especially rural and peri-urban areas in Kenya and Cambodia. In such contexts, limited digital infrastructure at schools, unstable conditions of the Internet Network, and insufficient access to devices for educational purposes pose significant barriers toward students' participation in digitally mediated teaching-learning environments (UNESCO; Trucano). As a result, students in these areas are subject to structured inequities in educational possibility (van Dijk), compared to students attending other schools with greater digital capacity. In this context, digital accessibility at the school level goes beyond device availability. It includes internet connectivity reliability in schools, along with functioning digital equipment for instructional use, and classroom preparation to support effective learning digitally (van Dijk; UNESCO). An initial number of devices must be shared between numerous pupils in many low-resource schools, and network breakdowns regularly interrupt instructional processes (Trucano; Hohlfeld et al.). These conditions interrupt the continuity of classroom instruction and decrease overall learning effectiveness.

Furthermore, constraints in digital accessibility at the school level may greatly impact students' psychological experiences. Frequent technical glitches, suboptimal learning environments, and little chance to interact with teachers and fellow students can frustrate, demotivate, and distance teens from the learning experience (Borup et al.; Ryan and Deci). This phenomenon, also referred to as psychological isolation, or students' perceived disconnection from the learning community, decreased willingness to participate, and lowered confidence in academic tasks, can have detrimental effects on motivation and retention. These effects are particularly severe among younger learners, who rely on structured interaction and feedback (Ryan and Deci; Borup et al.). Yet much of the digital divide literature has dealt primarily with macro-level factors such as national capacity/capability or system-wide policy initiatives and provided disproportionately less attention to the role school-level digital environments play in mediating students' academic performance and psychological functioning (van Dijk; UNESCO). Although these studies provide valuable insights, they often overlook the basic operational aspect of institutional learning environments offered in a constrained digital environment. More specifically, there is a notable lack of empirical research addressing the impact of limited school-based digital accessibility for elementary-age students in low-resource contexts (Trucano; UNESCO).

This study aims to fill this gap by focusing on elementary schools in remote and peri-urban areas of Kenya and Cambodia, where internet access is limited. This research is based on the actual implementation of online tutoring volunteer programs that these schools have conducted. Using survey data collected from participating students and teachers, this study also aims to explore the systematic effect of digital accessibility in school on academic participation and psychological isolation by examining inter-school differences in its utilization. By emphasizing school-based digital boundaries, this study offers a more nuanced understanding of the production of educational inequalities in institutional learning contexts. These findings underscore the need to reinforce digital infrastructure at the school level and foster a better learning environment in classrooms (van Dijk; UNESCO). Such findings contribute to practical understandings and policy insights that can inform the design of more effective, context-sensitive education interventions in low-resource settings.

2. Literature Review

2.1 School-Level Digital Accessibility in Education

School-level digital accessibility is defined by the availability of such an essential infrastructure in terms of access to stable internet connectivity, a sufficient number of digital devices, and adequately functioning classroom technology, which would enable learning. School-level digital-resource disparities drive education inequities, particularly in resource-poor settings (van Dijk; UNESCO; Trucano; Selwyn). Schools have to deal with various structural barriers like poor infrastructure, low budgets, and limited access to technological equipment in many of the developing regions (Trucano; UNESCO; van Dijk). Such constraints reduce

students' opportunities to engage with digital learning tools, which can ultimately hinder the effectiveness of instructional delivery. The quality of interaction with educational content is impaired, for instance, when digital devices are in limited supply or shared among multiple students (Hohlfeld et al.; Trucano; UNESCO).

However, much of the available literature is often based around broad measures of access such as national-level infrastructure or general technology uptake in schools (van Dijk; UNESCO; Trucano). This view misses out on the differences in access at the individual school level, where wide variances in infrastructure and resource allocation can help or hinder students' learning experiences. The above shows that the school-level digital context needs to be more localized for a better understanding of educational inequality in resource-constrained settings (van Dijk; UNESCO; Trucano).

2.2 Digital Learning Environments in Low-Resource Schools

Digital learning environments in low-resource schools are shaped by both technology capacity and institution dynamics. Ironically, many schools worldwide are faced with unreliable internet connections, aging devices, and a lack of trained personnel capable of bringing technology into the classroom in productive ways (UNESCO; Trucano; van Dijk; World Bank). As a result, digital tools are often underused or inconsistently applied, limiting their potential effect on student learning (UNESCO; Trucano; Hohlfeld et al.).

Studies have shown that the effectiveness of digital learning is intimately linked to the grade of the school climate, involving infrastructure, instructor preparedness, and classroom management practices (Hohlfeld et al.; Borup et al.; UNESCO). In schools lacking these vital components, digital learning initiatives frequently fail to yield significant gains in academic performance. In contrast, schools with stronger digital infrastructure and support systems have more success (Borup et al.; Hohlfeld et al.; UNESCO).

However, the current literature often assumes a degree of foundational digital operability that is not descriptive of the context in which low-resource environments are faced (UNESCO; Trucano; van Dijk). In such contexts, learners have fragmented learning experiences due to constant tech glitches, lack of access to devices, and ineffective support systems. Consequently, in order to understand the role of digital learning for low-resource schools, we need to determine the specialized institutional limitations that shape its implementation (Trucano; UNESCO; van Dijk).

2.3 Psychological Implications of Limited School-Based Digital Access

Low levels of digital accessibility in schools not only shape academic participation but also deeply influence students' psychological experiences. In digitally constrained school environments, ongoing technical difficulties and inconsistent access to learning tools can contribute to student frustration, disengagement, and diminished confidence (Borup et al.; Ryan and Deci; UNESCO).

Psychological isolation, in this context, is students' perceived detachment from classroom processes and their decreased readiness to connect with both teachers and peers. Research based in motivational theory shows that students require constant engagement and connection to the learning environment (Ryan and Deci; Borup et al.; UNESCO; Moore). When school-based digital environments do not support these interactions, the consequence can be decreased motivation and a sense of separation from learning (Ryan and Deci; Borup et al.; UNESCO).

However, much of the current literature examining psychological experiences in digital learning environments has been rooted in relatively well-resourced educational contexts (Borup et al.; UNESCO; van Dijk). Therefore, little is known about the ways that psychological isolation behaves in schools where access to devices is extremely limited. This gap is quite important since institutional constraints, such as resource-inconsistent infrastructure and sparse classroom interaction, could worsen the psychological burden among learners in low-resource conditions (Borup et al.; Ryan and Deci; UNESCO).

2.4 Research Gap and Study Contribution

Despite an abundance of literature exploring the digital divide, digital learning spaces, and student engagement, there still exist notable gaps. In the first place, a considerable part of the current literature focuses on macro-level analyses pertaining to national infrastructure or patterns in technology adoption at an aggregate level and therefore pays much less attention to school-level digital conditions (van Dijk; UNESCO; Trucano). Second, there is little empirical research on how constrained digital environments in schools can affect academic participation and psychological outcomes for elementary students in low-resource contexts (UNESCO; Trucano; Borup et al.).

In addition, studies dissecting these relationships during volunteer-based educational programs — where institutional resources can be scarce, and learning contexts differ widely (Trucano; UNESCO; Borup et al.) — are rare. This gap is particularly striking regarding the cross-country intersection of developing regions, where disparities in school infrastructure and success factors may manifest in very different educational outcomes (UNESCO; van Dijk; Trucano).

This study aims to fill this gap with a novel analysis of the ways in which differences in school-level access to digital technology affect academic participation and psychological isolation among elementary students from rural and peri-urban schools in Kenya and Cambodia. Drawing on real-life online tutoring volunteer programs, this study uses qualitative survey data from students and teachers to offer a more context-rich analysis of digital learning environments.

By examining issues of digital constraint, this study contributes to a better understanding of how educational inequalities are generated at the institutional level in schools. It additionally underscores the need for strengthening digital infrastructure at the school level and provides lessons learned in terms of designing effective and context-appropriate educational back-end system-based interventions suited to low-resource settings.

3. Theoretical Background

3.1 Digital Divide Theory (Institutional Perspective)

Digital Divide Theory is used to describe inequalities in access to digital technologies, such as the internet or mobile devices, and stresses the impact these gaps have on social and educational outcomes (van Dijk). Early conceptualizations of the digital divide largely centered on technology and Internet access, while newer treatments use an understanding that there are multiple layers of inequality, from access to usage to skills to context (van Dijk; UNESCO). In the educational sphere, the digital divide can be understood at different levels: national, institutional, and individual. The study borrows from an institutional perspective — examining digital accessibility specifically at the school-level. In low-resource contexts, schools are often disproportionately equipped with digital hardware and software for teaching and learning, leading to considerable gaps in student access to technology-enhanced learning environments (Trucano; UNESCO). These inequities can, in fact, hinder students' access to digital learning tools and thus the effectiveness of instructional delivery.

Moreover, Digital Divide Theory postulates that access to digital resources is paradoxically more likely to reinforce than reduce educational inequalities (van Dijk). Under-resourced education-based institutions may suffer from fragmented learning, less time spent engaging with educational material, or lower levels of academic involvement. This underscores the need to consider how institutional digital constraints affect educational outcomes in resource-poor contexts.

3.2 Social and Motivational Theory in Learning Environments

Social and motivation theories suggest that interaction, presence, and community are critical elements of successful learning environments. Self-Determination Theory posits that autonomy, competence, and relatedness significantly impact student motivation during the learning process (Ryan and Deci). Such factors are critical in fostering engagement and schema maintenance through positive educational outcomes. In learning environments based in schools, teacher and peer interaction is key to students' motivation and participation (Borup et al.; Ryan and Deci). Yet when digital learning contexts are limited by inadequate infrastructure, meaningful engagement can also be diminished. For instance, unstable internet connections or lean technology in the classroom can cause a communication bottleneck and hamper students' active engagement during lessons. It implicates school-based digital accessibility as not just a technical matter but one directly impacting students' psychological engagement. Progression to these interactions being impeded can lead to students feeling less invested in the classroom atmosphere, having lower confidence levels, or a lack of belonging. These results are especially important for young learners, who depend on structured interaction and immediate feedback.

3.3 Psychological Isolation in Digitally Constrained School Environments

Psychological isolation is a condition in which students feel alienated from their context for learning, peers, and instructors. In the context of school systems that are digitally constrained, this is likely a result of repeated technical disruptions, inconsistent access to learning tools, and limited opportunities for interaction (Borup et al.; UNESCO). In the classic classroom, students find themselves well-positioned for interaction due to their physical presence; in contrast, digitally mediated environments are heavily reliant on technological systems and methods. When these systems are weak or unreliable, they put up barriers to students participating and can lead them to feel disengaged or disconnected from what is happening in class.

Research also shows that psychological isolation is strongly associated with academic results. Students who feel disengaged from the sociocultural learning environment are less likely to contribute, complete assignments, and ask for help when necessary (Ryan and Deci; Borup et al.). In light of this, institutional constraints in underserved school settings may further compound these effects, thus making psychological isolation an integral aspect of the understanding of education inequality more broadly. Psychological isolation is conceptualized in the present study as a crucial outcome variable affected by school-level digital accessibility. This study offers a more holistic view of how digital inequality functions within the school context by linking institutional-level digital barriers to students' psychological experiences.

4. Research Method

4.1 Research Design

Using a mixed-method approach leveraging quantitative survey data and qualitative observational context, this study investigates the relationship between the level of school digital accessibility and participation in academic folk groups as well as psychological isolation. The study is based on practical educational work happening in the real world, namely online tutoring volunteer programs, which have been used since early 2020 in elementary schools in rural and peri-urban areas of Kenya and Cambodia.

This study will synthesize multiple data sources in order to offer a fuller understanding of the ways institutional digital conditions shape students' academic as well as psychological learning experiences. It enables triangulation – meaning the combination of both qualitative and quantitative approaches provides an unprecedented basis on which to analyze and validate findings.

4.2 Participants

In this study, the participants are elementary school students and teachers who volunteer to participate in online tutoring programs at specific schools in Kenya and Cambodia. The sample consists of approximately 80–200 students from several participating schools, as well as a smaller group of teachers and local coordinators who were involved in program delivery. The study invites students who are predominantly in lower elementary grades who require significant guidance and interaction for knowledge assimilation. Educators and coordinators offer unique

insights into classroom dynamics, student engagement, and other challenges around digital accessibility.

4.3 Data Collection

Data was collected through structured surveys administered to both students and teachers, supplemented by observational records from the online tutoring sessions. The surveys were designed to capture three main dimensions:

- **School-Level Digital Accessibility**
(e.g., availability of devices in classrooms, reliability of school internet connectivity, frequency of technical disruptions)
- **Academic Participation**
(e.g., class attendance, level of engagement during lessons, completion of assigned tasks)
- **Psychological Isolation**
(e.g., feelings of disconnection, reluctance to participate, confidence in asking questions)

Survey questions were adapted to be age-appropriate for elementary students and were administered with the assistance of local coordinators where necessary. In addition, observational notes recorded by volunteer tutors were used to capture contextual factors such as classroom interaction patterns and technical issues during sessions. The full survey instruments used in this study are provided in Appendix A and Appendix B. All survey items were measured using a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree).

4.4 Variables and Measurement

This study examines the relationship between one independent variable and two dependent variables:

- **Independent Variable:**
School-Level Digital Accessibility
→ measured through indicators such as device availability, network stability, and frequency of technical interruptions within the school environment
- **Dependent Variables:**
 1. Academic Participation
→ measured through attendance rates, participation frequency, and task completion
 2. Psychological Isolation
→ measured through self-reported feelings of disconnection, reduced interaction, and lack of confidence in engagement

Each variable was assessed using a combination of survey responses and observational data to improve measurement reliability.

4.5 Data Analysis

Descriptive and comparative statistical methods were used for data analytics. Initially, frequency analysis was conducted to inspect broad-level tendencies in digital-led availability, contribution, and mental states of individuals. Second, a comparative analysis was undertaken to examine differences in academic engagement and psychological isolation at different levels of school-level digital accessibility. Moreover, simple correlations were used to examine the relationships between digital accessibility and the two outcome variables. This method also enables the identification of common trends without over-inflating causal connections because of the exploratory design of this study. Qualitative observational data were applied to contextualize and substantiate the quantitative findings, so as to provide further insight into how digital constraints play out in practice.

5. Results

5.1 School-Level Digital Accessibility

Survey results show school-level digital accessibility is limited in participating schools across Kenya and Cambodia. Many students explained that they do not have access to adequate amounts of digital devices and, many times, said devices must be shared among users. Also, a stable internet connection has repeatedly proved to be a major barrier for both students and teachers in the effective learning process.

Indeed, about 65% of student respondents reported incessant interruptions when utilizing online tutors due to technical difficulties. This observation was reinforced by teacher responses, with over 70% stating that lessons were often disturbed because of connectivity issues. Such results point to school-level Digital Infrastructure as an important constraint on the delivery of online learning in low-resource settings.

5.2 Academic Participation

Student participation is clearly linked to digital accessibility, per analysis of respective survey responses. Those attending schools with better digital conditions reported higher rates of attendance and engagement, including up to 68% participation, compared to much lower rates in more constrained digital settings. In contrast, student participation was reduced for those experiencing frequent technical interruptions.

In other words, students with reliable internet access were more likely than others to participate during lessons and complete assignments consistently. These trends were confirmed by teacher observations, which found that levels of participation tended to decrease in classrooms with more technical problems. These findings indicate that school-level digital accessibility has strong implications for students' ability to participate in learning activities.

5.3 Psychological Isolation

Survey findings also reveal notable patterns related to students' psychological experiences. Students in digitally constrained school environments reported higher levels of frustration and reduced confidence in participating during lessons. A significant portion of respondents indicated that they felt hesitant to ask questions or participate actively when technical issues occurred.

Furthermore, students who experienced frequent disruptions were more likely to report feelings of disconnection from their teacher and classmates. Teacher responses supported this trend, with many noting that students in unstable digital environments appeared less engaged and more withdrawn during lessons. These findings suggest that limitations in school-level digital accessibility contribute not only to reduced academic participation but also to increased psychological isolation among students.

5.4 Relationship Between Digital Accessibility, Participation, and Isolation

Correlation analysis indicates a consistent relationship between school-level digital accessibility and both academic and psychological outcomes. Higher levels of digital accessibility were associated with increased student participation and lower levels of psychological isolation. Conversely, limited digital access was linked to decreased engagement and stronger feelings of disconnection.

Although the analysis does not establish causality, the observed patterns suggest that school-level digital conditions are closely associated with both students' learning behaviors and their psychological experiences. These results highlight the importance of improving digital infrastructure at the school level to enhance both academic engagement and student well-being.

6. Discussion

This study highlights the importance of school-level digital accessibility that can determine academic socialization as well as psychological experiences for elementary students in low-resource contexts. However, in participating schools in Kenya and Cambodia, poor digital infrastructure and poor internet connectivity were consistently associated with lower rates of student engagement and increased psychological isolation. These findings buttress the central argument of Digital Divide Theory, contending that unequal digital access can potentially magnify alternative educational inequalities, instead of relieving them (van Dijk). Working at the micro-macro level of educational access, this study illustrated how school-level differences in digital infrastructure translated into different experiences for students trying to engage with learning activities. Schools with relatively steady digital behavior tended to have greater attendance and engagement and more tasks completed, while schools that experienced frequent technical breakdowns saw this participation wane.

Specifically, beyond academic outcomes, this study adds to the literature that highlights the international significance of digital accessibility's impact on students' psychological well-being. In accordance with Social and Motivational Theory, diminished opportunities for

interaction and repeated technical obstacles were correlated with a lack of motivation as well as lower perceptions of belonging among students (Ryan and Deci; Borup et al.). This suggests that school-level digital conditions impact not just what students learn, but how they experience that learning as well. Moreover, the results highlight that digital accessibility at a school relates to systemic issues rather than individual factors. While previous research has often focused on national infrastructure or aggregate access trends, this study shows that the conditions present in individual schools are a major factor determining educational experiences. Such an institutional perspective is particularly appropriate in low-resource settings, where there can be wide variation between schools.

From a policy perspective, these findings illustrate the significance of enhancing school-level digital infrastructure before we can think about data literacy. First, internet connectivity in schools: the need of the hour is digital continuity in learning, and for this, setting up a reliable internet connectivity system within schools is an all-time basic. Second, expanding access to classroom-based digital devices can help reduce the need for device sharing and promote more productive student engagement with learning materials. Third, teacher training and support around effective use of digital tools in instructional practices will help address both engagement and learning outcomes. In addition, the psychological dimension that we found in this study indicates that policymakers on digital infrastructure cannot only focus their enhancements on technical aspects. Similarly, supporting students' motivation and emotional well-being with stable, interactive learning environments is vitally important. This includes building classroom practices that welcome participation, as well as in technologically constrained environments.

Finally, this study has some potential limitations that must be stated. Since the data rely on self-reported surveys, there may be biases, especially among younger students. In addition, the sample is limited to purposefully selected schools in Kenya and Cambodia, which may affect generalizability. Future studies could widen the sample by using more schools and a far larger dataset for longitudinal modelling, thus providing insight into causal relationships. Despite these limitations, this study provides important insights about how school-level digital accessibility relates to academic and psychosocial outcomes. Emphasizing the role of institutional digital conditions, the findings give context for ongoing debates around educational equity and provide grounding for more impactful and context-sensitive policy interventions in low-resource settings.

7. Conclusion

This study advances existing research by showing that school-level digital accessibility is an important predictor of access to, and experience with, academics for elementary students in under-resourced schools. The findings provide empirical evidence of how limited digital infrastructure and unstable internet connectivity in schools deter uptake for learning activities, with the focus on Kenya and Cambodia as case studies. This study grounds discussions of digital inequality at an institutional level, expanding macro-level conversations about the digital divide and highlighting how gaps in school-based digital environments yield observable differences in

attendance, engagement, and task completion. In doing so, it contributes to understanding the reproduction of educational inequalities within daily learning situations.

Importantly, this study adds to the literature by introducing multivariate psychological dimensions into the analysis of digital accessibility. However, the irony of technology-based learning and the findings illustrate that constant tech disruptions with a lack of interaction not only pose an impediment to academic engagement but also leave dwindling motivation, shaken confidence, and psychological pain. This highlights the two-pronged effect of digital inequality on learning outcomes and student well-being. Connecting structural digital constraints to the subjective learning experiences of students, this study extends our understanding of how the digital divide functions in both cognitive and affective domains. This integrative perspective constitutes a valuable contribution to one of the more consequential conversations happening in education research today, especially for low-resource contexts.

From a policy and practice perspective, the results demonstrate the importance of investing in school-level digital infrastructure for equitable education. As well, developing infrastructure to ensure reliable internet connectivity and visible, supportive classroom environments over digital devices that are accessible and of adequate quality for ensuring meaningful access in digitally mediated learning. Moreover, the capacity for teachers to be able to effectively adapt digital tools into their pedagogy is crucial in ensuring that investments in infrastructure lead to improved learning outcomes. Although its contention to self-reported data limits the study, and the reach of the sample is unequal in terms of geographical representation, it provides a strong basis for future studies involving bigger datasets and longitudinal methods to prove relations of cause. In summary, this study adds substantial value by emphasizing the importance of institutional digital conditions in affecting both academic and psychosocial outcomes, as well as presenting actionable implications for designing more equity-oriented and contextual educational interventions in low-resource settings.

Works Cited

- Borup, Jered, et al. "Academic Communities of Engagement: An Expanded Framework for K–12 Online Learning." *Educational Technology Research and Development*, vol. 68, no. 2, 2020, pp. 807–832.
- Hohlfeld, Tina N., et al. "Examining the Digital Divide in K–12 Public Schools: Four-Year Trends for Supporting ICT Literacy in Florida." *Computers & Education*, 2017.
- Moore, Michael G. "Three Types of Interaction." *American Journal of Distance Education*, vol. 3, no. 2, 1989, pp. 1–7.
- Ryan, Richard M., and Edward L. Deci. "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions." *Contemporary Educational Psychology*, vol. 25, no. 1, 2000, pp. 54–67.
- Selwyn, Neil. *Education and Technology: Key Issues and Debates*. Bloomsbury Academic, 2016.
- Trucano, Michael. *Knowledge Maps: ICT in Education*. World Bank, 2016.
- UNESCO. *Education in a Post-COVID World: Nine Ideas for Public Action*. UNESCO, 2020.
- van Dijk, Jan. *The Digital Divide*. Polity Press, 2020.
- World Bank. *World Development Report 2016: Digital Dividends*. World Bank, 2016.

Impact of Maternal Age and Body Mass Index (BMI) on IVF success By Summer Shum

Abstract

In vitro fertilisation (IVF) is a highly prominent and advanced treatment used for infertility that has modernised the field of assisted reproductive technologies. However, as with most treatments, numerous factors influence the success of IVF. Age and body mass index (BMI) are two key determinants of IVF success. This research paper aimed to evaluate the contribution of these two factors on IVF success rates in a clinical cohort. We analyzed a cohort of 2,232 women from an IVF dataset available on the Harvard Dataverse. This provided us with clinical and demographic data on age and BMI, with IVF success serving as the primary outcome. Overall, we found that IVF success declined with advancing age. To further support the strong association between age and IVF success, we tested whether egg source influenced IVF success and found that oocytes derived from donors had a higher IVF success rate than those from patients using their own (autologous) oocytes. This observation is attributed to the younger age of donors with better oocyte quality. Likewise, BMI and IVF success shared a negative correlation, indicating that a higher BMI is associated with lower IVF success rates. Our findings would help IVF clinics or policymakers refine the criteria selection for subsidised treatments. Our paper also highlights the role of risk factors affecting IVF success rates, allowing doctors to individualise treatment plans for patients to ensure the best possible outcome.

Introduction

Infertility represents a profound global clinical and psychosocial challenge that affects 10% to 15% of reproductive-age couples (Mohamed Salih et al.). In vitro fertilisation (IVF) is a type of assisted reproductive technology (ART) that consists of a complex series of medical procedures that were designed to achieve conception in the late 1970s. Since the introduction of IVF to the field of ART, IVF has revolutionized the clinical management of infertility. To this day, IVF remains the most widely used treatment approach in this field (Mohamed Salih et al.). The development of IVF not only established a new field in ART (Elizur et al.) but also expanded reproductive treatments for infertile couples by addressing fertility issues in both sexes, such as intracytoplasmic sperm injection (ICSI). ICSI specifically addresses severe male-factor infertility (Ribeiro and Sousa), thereby further broadening the field of ART.

IVF treatment comprises complex clinical and laboratory procedures that result in the formation of an embryo. Successful implantation of the embryo into the uterine lining during transfer results in pregnancy (Mizrachi and McQueen). These procedures include controlled ovarian stimulation, retrieval of oocytes, fertilisation outside the body, and the eventual transfer of embryos into the uterus (Mohamed Salih et al.).

The success of IVF treatment is quantified by the crude pregnancy rate, defined as the number of deliveries per treatment cycle (Elizur et al.). This clinical data has been used by researchers to develop predictive models by utilizing conventional statistical methodologies to

estimate each individual's probability of IVF success. These success-rate-based models help physicians tailor their treatment strategies based on prior clinical outcomes (Shingshetty et al.).

Despite significant advancements in IVF technology and embryology, the success rate of IVF remains variable and is still heavily influenced by biological, environmental, and clinical factors. These factors include maternal age and body mass index (BMI). Careful consideration and management of these variables can optimize the success of an IVF cycle. The goal of this paper was to evaluate how each factor independently affects IVF success outcomes.

In recent decades, the number of women who prioritize their education, careers, and financial goals over settling down early to start a family has been rising (Li et al.). However, advancing maternal age is linked to a reduced ovarian reserve, hence resulting in a higher prevalence of infertility (Gao et al.). The Ovarian reserve (OR) primarily measures the quantity of the remaining number of follicles, which is another predictor of IVF success (Broekmans et al.). Ribeiro & Sousa reviewed the effects of female age and ovarian reserves on IVF outcomes (Ribeiro and Sousa). Their results showed that increasing female age decreased the mean number of retrieved oocytes, the embryo cleavage rate, and the biochemical pregnancy rate. These all correlated with poorer cycles, higher cancellation and miscarriage rates, and overall, reduced IVF success.

Additionally, advanced maternal age is clinically associated with an impaired response to controlled ovarian hyperstimulation (COH), resulting in lower number of eggs retrieved for embryogenesis, which fundamentally constrains IVF success (Dabbagh Rezaeiye et al.).

Oocyte quality is another key factor that influences the rate of successful fertilization and early embryogenesis. According to a study by Kenney and McGowan, there were 18,530 ART cycles in the US involving egg donors in 2011 (Kenney and McGowan). This represents a significant rise in utilisation to help achieve pregnancies from better oocyte quality since the first successful application of this technology in 1983 (Kenney and McGowan).

Nonetheless, IVF success rates remain influenced by similar factors as those observed in cycles using autologous oocytes, on both the recipient and the donor. These include the donor's age, the recipient's ovarian health, and whether the eggs were fresh or previously vitrified (Martinez et al.). Using the implantation rate serves as an indicator of the recipient's uterine receptivity and provides the essential foundation for determining the effect of Ova source on IVF success (Gupta et al.). Apart from the biological quality of the oocyte, there are other administrative and ethical guidelines surrounding ova donation, such as donor reimbursement for time and discomfort. This limits the number of high-quality eggs available for donation and, hence, also plays a role in determining the IVF success (Kenney and McGowan). This paper used the rate of IVF success using Ova donors as key evidence to further support the impact of age on IVF success.

BMI is calculated as weight in kilograms divided by height squared in metres and is used to estimate body fat (Liu et al.). Obesity is characterised by excess body fat and is highly regarded as a factor of numerous health conditions (Penzias et al.). Early on-set obesity (having a $BMI \geq 30.0$) in women predisposes to menstrual irregularities, chronic oligo-anovulation, and

infertility in adulthood. Therefore, a woman with a BMI ≥ 30.0 has an increased chance of miscarriage and an overall reduced success of pregnancy outcomes. This is due to metabolic and endocrine imbalances within their uterine environment (Boddeti et al.). Furthermore, obesity negatively affects oocyte quality and endometrial responsiveness (Boddeti et al.).

This study consequently evaluates these two factors - age and BMI in relation to IVF success prevalence within a cross-sectional cohort. Crude pregnancy rates were used as a measure of IVF success rate. We hypothesized that age and BMI would correlate inversely with IVF success rate, that ova source would demonstrate differential success rates reflecting biological and hormonal influences on the donor. By combining factors such as metabolic health, ovarian ageing, and egg source in a systematic evaluation, we aim to clarify the relative contributions of each factor to IVF success within the presented cohort. We analyzed patterns in the dataset to understand how each factor influenced overall reproductive outcomes.

Method

Data were obtained from an IVF dataset hosted on Harvard Dataverse, accessible at <https://doi.org/10.7910/DVN/GTBE1A>. The dataset includes patient demographics, clinical measurements, and IVF outcomes.

All analyses were performed in R, utilizing standard packages such as “ggplot2” and “dplyr” for data cleaning, normalization, visualization, and statistical tests. Descriptive statistics and visualizations were used to summarize the distributions of age, BMI, and ova source in relation to IVF success outcomes. Relationships between continuous variables were assessed using Spearman correlation tests, while significance in variable distributions was determined by Wilcoxon rank-sum tests. Chi-square tests were used to evaluate associations between categorical variables. Specific cutoffs for cohort separation, such as distinguishing between old and young groups, were selected based on clinical relevance (Faddy et al.). Results with p-values below 0.05 were considered statistically significant.

Results

Age and IVF outcomes

Maternal age is widely recognised as one of the most critical determinants of success in IVF (Cimadomo et al.). Hence, we evaluated whether a woman's age reduces her overall probability of achieving a successful IVF outcome. To investigate the relationship between maternal age and IVF success, we first coded IVF outcomes as binary, with 1 indicating success and 0 indicating failure. The IVF success rate was calculated for each age by computing the mean binary IVF outcome against age (Figure 1). First, we investigated whether there was a correlation between patients' age and IVF success rate.

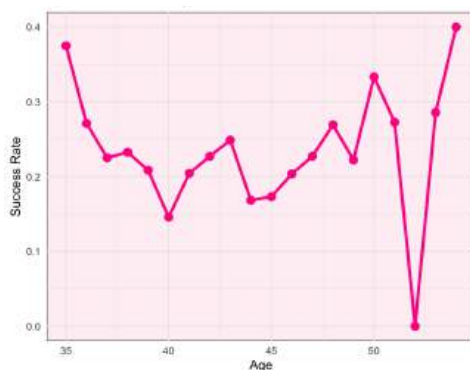


Figure 1. Age vs IVF Success rate

Our data demonstrated a general decline in IVF success rate with increasing age, indicating a negligible negative correlation between age and IVF success rate, but this is not statistically significant (Spearman's rank correlation; $r = -0.01$, $p = 0.68$) (Figure 1).

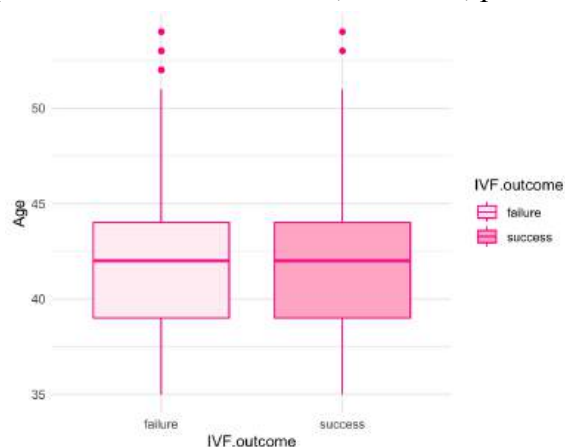


Figure 2. Comparison of the mean age of each success and failure group

In addition, we compared the mean ages of the IVF success and failure groups and found that the mean ages were similar, indicating no statistically significant difference (Wilcoxon rank-sum test with continuity correction; p -value = 0.68) (Figure 2). This suggests that, although an almost negligible downward trend is evident, age alone does not show a strong, continuous association with IVF success across all ages. Hence, to analyse this further, patients were divided into 2 age groups: Younger (<37 years old) and older (≥ 37 years old) (Figure 3). We followed a study by Faddy et al., which found that the rate of follicle loss is constant until age 37, at which point it doubles, using 37 years as our cutoff age for each group (Faddy et al.). For each group, the total number of IVF cycles and the number of successful outcomes were calculated.

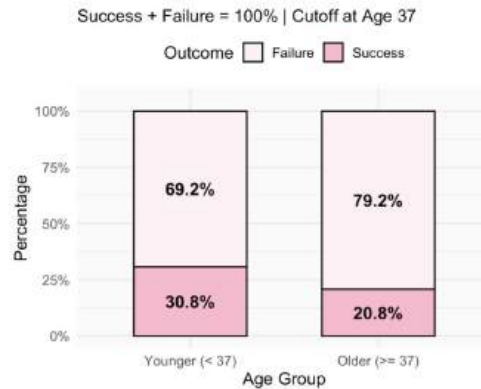


Figure 3. Comparison of IVF success ratios between Younger (<37 years old) and Older ($\geq$ 37 years old) groups

A comparison of the rate of IVF outcomes showed that younger individuals below the age of 37 exhibited a statistically significantly greater prevalence of IVF success compared to older individuals above the age of 37, approximately 10% greater (IVF success rate of younger group: 30.8%, older group: 20.8%; Chi-square test ; $p = 0.03$) (Figure 3). This finding may support the idea that, as age increases, both the quality and number of oocytes available naturally decline, which in turn reduces the probability of achieving a successful IVF outcome.

To further support the idea that age plays a critical role in IVF success rates, we also tested whether there were differences in the IVF success between donor ova and patient ova. Preceding research suggests that the probability of IVF success differs between cycles using Donor ova and those using patient (autologous) ova (Yeh et al.). To evaluate the association between Egg source and IVF success rate, we first assigned the value “1” to represent a successful IVF outcome and “0” to represent an unsuccessful/failed IVF outcome. We analysed the number of successes and failures in each group, Donor and Patient ova, before converting them into percentages of their total (Figure 4).

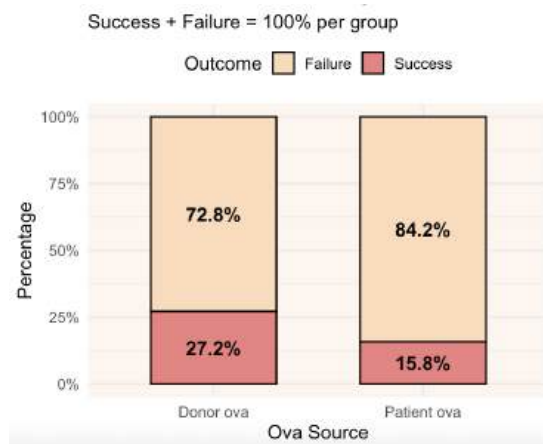


Figure 4. Outcome distribution by ova source. Graph showing % of Success and Failure in each group

Our results showed that there were more successful IVF outcomes among those using a Donor (27.2%) than among those using their own ova (15.8%). Likewise, those using their own

ova had a higher rate (84.2%) of failed IVF outcomes compared to those who used a donor (72.8%). Hence, our findings suggest the use of Donor ova has a statistically significantly greater IVF success outcome (Chi-square test ; $p = 7.72 \times 10^{-11}$) (Figure 4).

BMI and IVF outcomes

Based on prior research, having an elevated BMI is associated with hormonal dysregulation that impairs normal ovulation and overall ovarian function (Snider and Wood). Hence, patients with higher BMI have fewer and poorer-quality oocytes retrieved, which can subsequently compromise embryo development prior to implantation. This results in a lesser number of good-quality pregnancy potential embryos as well as an increased risk of miscarriage, ultimately lowering overall IVF success rates (Shah et al.). Consequently, investigating the relationship between BMI and IVF outcomes may clarify the extent to which metabolic status influences reproductive potential. We hypothesized that BMI and IVF success rate would be negatively correlated. To evaluate the relationship between BMI and IVF success rate, we first assigned the value “1” to represent a successful IVF outcome and “0” to represent an unsuccessful/failed IVF outcome. We analysed the IVF success rate as a function of BMI (Figure 5).

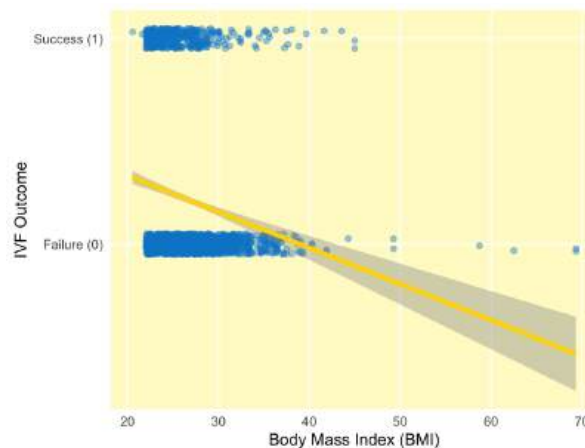


Figure 5. The correlation between IVF success rate and BMI.

The results demonstrated a statistically significant negative correlation between BMI and IVF outcome, indicating that IVF success rate declines with increasing BMI (Spearman's rank correlation ; $r = -0.23$, $p = 5.70 \times 10^{-29}$) (Figure 5).

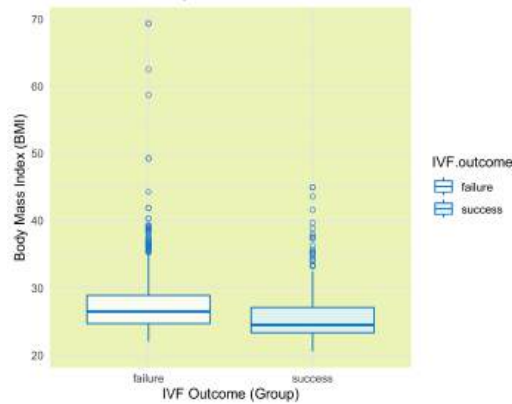


Figure 6. Distribution of BMI and IVF outcome

In addition, we compared the mean BMI of the IVF success and the failure groups (Figure 6) and found that the mean BMI of the success group (Mean_BMI: 25.58) is statistically significantly lower than that of the failure group (Mean_BMI: 27.22) (Wilcoxon rank-sum test with continuity correction; $p\text{-value} = 3.08 \times 10^{-28}$) (Figure 6). Hence, our data confirm an inverse association between BMI and IVF success, with higher BMI associated with lower IVF success rates. Moreover, to further investigate the impact of BMI on IVF success rate, subjects were categorized by BMI into 2 groups: those with a BMI below 30 (Normal) and those with a BMI above 30 (Overweight) (Figure 7). We followed the World Health Organisation (WHO) BMI classification system to determine those who had a normal or overweight BMI (*Body Mass Index (BMI)*).

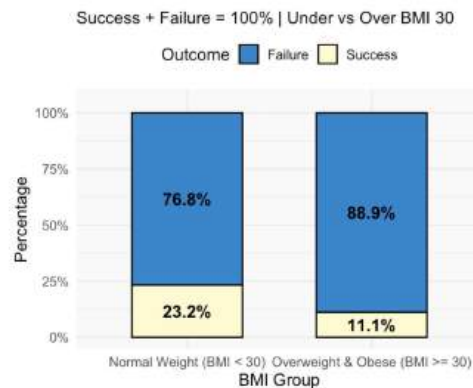


Figure 7. Comparison of IVF success ratios between Normal weight (BMI <30) and Overweight (BMI ≥30) groups

We compared the mean percentage success rates of the two groups (Figure 7). In this study cohort, there were 2232 subjects; the vast majority ($n = 1863$) were in the Normal BMI group, while approximately one-sixth ($n = 369$) were classified as overweight. The average age of each sample was consistent between both groups, providing a stable baseline for comparison (Age = 42). Data analysis revealed a noticeable difference in the mean IVF success rates between the two groups. Patients in the Normal (<30) group achieved a higher mean success rate of 23.2%, while those in the Overweight (≥ 30) group achieved a mean success rate of 11.1%. This

indicates a statistically significant association between higher IVF success and lower BMI, supporting previous evidence of an inverse relationship between BMI and IVF success (Chi-square test; $p = 3.11 \times 10^{-7}$) (Figure 7).

Discussion

The study examined the relationship between age and BMI on IVF success. There were two main findings. First, IVF success rates decreased with advancing age. Although the mean age of successful and unsuccessful IVF outcomes is not significantly different, younger individuals (below 37) had significantly higher IVF success rates than older individuals (those 37 or older). Secondly, obese individuals with a BMI greater than 30 have a significantly lower number of IVF successes than those of normal weight and a BMI below 30, indicating that BMI would be another key factor for IVF success rate.

Maternal age remains one of the main determinants of IVF success (Cimadomo et al.). Not only does a woman's ovarian reserve decline with age, but physiological pathways such as energy production and balance, metabolism, epigenetic regulation, and the cell cycle during embryonic development are also altered, compromising the quality of oocytes and embryos (Cimadomo et al.). This combination of factors invariably leads to a reduced IVF success rate.

Ova source could be used as another major factor to support the evidence behind advancing age and the declining IVF success rate. Our data indicate donor ova from younger donors has a higher success rate. This is the reason the use of donor oocytes has gained popularity as a treatment for age-related decline in ovarian reserve (Robin et al.). It has been established that advancing maternal age is associated with declines in both oocyte quality and endometrial receptivity (Robin et al.). In general, most women who use donor ova are in their early or late forties. As expected with increasing age, there are physiological alterations in the uterine lining, including altered endometrial steroid receptor activity, altered blood flow, and stromal angiosclerosis, which negatively affect implantation rates (Robin et al.). Furthermore, the finding of an increased pregnancy rate among those using Donor eggs highlights the importance of ovarian reserve and egg quality for IVF pregnancy success (Robin et al.).

BMI is another key determinant of IVF success. Obesity is associated with a relative resistance to gonadotropin, a hormone that regulates ovarian function (Petanovski, Z et al.). Hence, patients with a higher BMI often receive higher gonadotropin doses and undergo extended ovarian stimulation, yet still produce lower serum estradiol concentrations and fewer oocytes, indicating an altered response to ovarian hyperstimulation (COH) (Petanovski, Z et al.). Additionally, hormonal imbalances, including elevated insulin and androgen levels, are observed in obese patients, further interfering with ovarian function and overall oocyte quality (Boddeti et al.). Obesity is also often expressed in women with polycystic ovary syndrome (PCOS) (Alenezi et al.). Women with this condition often have increased visceral subcutaneous body fat due to hyperinsulinemia and insulin resistance (Meneghini et al., 2023). Despite requiring an increased dose of FSH during ovarian stimulation, fewer oocytes are collected, and serum estradiol concentration decreases, contributing to higher cycle cancellation rates and ultimately impairing

pregnancy rates. Other significant implications of PCOS include reduced oocyte and embryo quality, abnormal endometrial development, and impaired implantation.

There were several limitations that should be taken into account. Firstly, it is unclear how a successful IVF outcome was defined in the dataset. IVF success may refer to biochemical pregnancy (Annan et al.), clinical pregnancy confirmed by ultrasound, ongoing pregnancy beyond the first trimester (Seungdamrong et al.), or live birth (Garrido et al.). Without a precise definition, the interpretation of reported IVF success becomes inconclusive. Secondly, our dataset lacks diversity as it was sourced from an entirely Mexican population. This limits the external validity of the study, as the results may not be representative of other ethnic groups due to genetic, environmental, and lifestyle differences across ethnic populations. Consequently, these differences may independently influence age, BMI, and ovarian source, ultimately affecting IVF success. Apart from ethnicity, other key variables such as day of embryo transfer, semen quality, previous pregnancy (Shingshetty et al.), and cause of infertility were not uniformly captured. This further limits the ability to fully account for the confounding factors on IVF success. Thirdly, while BMI cutoffs were used as an accessible clinical tool in this study, BMI does not account for other factors, such as frame size or an individual's lean body mass, potentially misclassifying those with high muscle density as overweight (Penzias et al.). Furthermore, BMI does not distinguish among fat distribution patterns, such as peripheral or abdominal fat, which may have distinct effects on IVF success and reproductive outcomes. Hence, the dataset fails to identify metabolic risks associated with body fat among Mexican individuals in relation to IVF success. There were also sample size limitations in this dataset, particularly the underrepresentation of women older than 37; there were only 91 out of the 2232 present individuals. This may have led to our finding that age had an insignificant negative correlation with IVF success, thereby further reducing the generalizability of the results.

Therefore, future studies should leverage longitudinal, population-representative datasets with comprehensive covariate capture to validate these findings. Finally, establishing a clear definition of IVF success, together with objective measures including maternal age, BMI, and ova source, can improve the assessment of IVF outcomes and support more personalized clinical decision-making.

By examining metabolic health, ovarian ageing, and egg source, the study's findings collectively suggest that these factors shape the success of IVF in the dataset. Although IVF success rates generally followed the expected pattern with age, the dataset showed some variation, suggesting that additional factors may be contributing to the results. Differences in IVF outcomes between donor and patient ova highlight the relative contributions of biological factors, such as oocyte quality and both maternal age and metabolic status, as reflected by BMI. Furthermore, BMI is a predictive indicator of IVF success, underscoring the value of integrating physiological, reproductive, and demographic factors when assessing reproductive potential. Our findings could serve as a baseline for hospitals in shaping and refining criteria for IVF treatments, aiming to make care more personalized and manage risk factors effectively to optimize IVF success.

Works cited

- Alenezi, Salih Atalah, et al. 'The Impact of High BMI on Pregnancy Outcomes and Complications in Women with PCOS Undergoing IVF—A Systematic Review and Meta-Analysis'. *Journal of Clinical Medicine*, vol. 13, no. 6, Jan. 2024, p. 1578. www.mdpi.com, <https://doi.org/10.3390/jcm13061578>.
- Annan, John Jude Kwelu, et al. 'Biochemical Pregnancy During Assisted Conception: A Little Bit Pregnant'. *Journal of Clinical Medicine Research*, vol. 5, no. 4, June 2013, pp. 269–74. COinS, <https://doi.org/10.4021/jocmr.v5i4.1008>.
- Boddeti, Sruthi, et al. 'Impact of Obesity on In Vitro Fertilization (IVF) Outcomes: A Systematic Review'. *Cureus*, vol. 17, no. 8, Aug. 2025. www.cureus.com, <https://doi.org/10.7759/cureus.90695>.
- Body Mass Index (BMI). <https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/body-mass-index>. Accessed 28 Jan. 2026.
- Broekmans, F. J., et al. 'A Systematic Review of Tests Predicting Ovarian Reserve and IVF Outcome'. *Human Reproduction Update*, vol. 12, no. 6, Dec. 2006, pp. 685–718. academic.oup.com, <https://doi.org/10.1093/humupd/dml034>.
- Cimadomo, Danilo, et al. 'Impact of Maternal Age on Oocyte and Embryo Competence'. *Frontiers in Endocrinology*, vol. 9, June 2018. *Frontiers*, <https://doi.org/10.3389/fendo.2018.00327>.
- Dabbagh Rezaeiye, Radin, et al. 'Impact of Various Parameters as Predictors of The Success Rate of In Vitro Fertilization'. *International Journal of Fertility & Sterility*, vol. 16, no. 2, 2022, pp. 76–84. PubMed Central, <https://doi.org/10.22074/IJFS.2021.531672.1134>.
- Elizur, Shai E., et al. 'Factors Predicting IVF Treatment Outcome: A Multivariate Analysis of 5310 Cycles'. *Reproductive BioMedicine Online*, vol. 10, no. 5, Jan. 2005, pp. 645–49. DOI.org (Crossref), [https://doi.org/10.1016/S1472-6483\(10\)61673-2](https://doi.org/10.1016/S1472-6483(10)61673-2).
- Faddy, M. J., et al. 'Accelerated Disappearance of Ovarian Follicles in Mid-Life: Implications for Forecasting Menopause'. *Human Reproduction*, vol. 7, no. 10, Nov. 1992, pp. 1342–46. DOI.org (Crossref), <https://doi.org/10.1093/oxfordjournals.humrep.a137570>.
- Gao, Jinyu, et al. 'Factors Influencing Clinical Pregnancy Outcome of in Vitro Fertilization/Intracytoplasmic Sperm Injection in Older Women'. *African Health Sciences*, vol. 23, no. 2, June 2023, pp. 632–39. PubMed Central, <https://doi.org/10.4314/ahs.v23i2.73>.
- Garrido, Nicolás, et al. 'Cumulative Live-Birth Rates per Total Number of Embryos Needed to Reach Newborn in Consecutive in Vitro Fertilization (IVF) Cycles: A New Approach to Measuring the Likelihood of IVF Success'. *Fertility and Sterility*, vol. 96, no. 1, July 2011, pp. 40–46. www.fertstert.org, <https://doi.org/10.1016/j.fertnstert.2011.05.008>.
- Gupta, Pooja, et al. 'A Study of Recipient Related Predictors of Success in Oocyte Donation Program'. *Journal of Human Reproductive Sciences*, vol. 5, no. 3, Dec. 2012, p. 252. journals.lww.com, <https://doi.org/10.4103/0974-1208.106336>.

- Kenney, Nancy J., and Michelle L. McGowan. 'Egg Donation Compensation: Ethical and Legal Challenges'. *Medicolegal and Bioethics*, vol. 4, Sept. 2014, pp. 15–24. Taylor and Francis+NEJM, <https://doi.org/10.2147/MB.S51328>.
- Li, Xin, et al. 'Postponed Childbearing: A Cross-Sectional Study of Differences between Subjective and Objective Factors'. *Annals of Medicine*, vol. 57, no. 1, p. 2546674. PubMed Central, <https://doi.org/10.1080/07853890.2025.2546674>.
- Liu, Dan, et al. 'Effects of Body Mass Index on IVF Outcomes in Different Age Groups'. *BMC Women's Health*, vol. 23, no. 1, Aug. 2023, p. 416. Springer Link, <https://doi.org/10.1186/s12905-023-02540-8>.
- Martinez, Francisca, et al. 'Ovarian Stimulation for Oocyte Donation: A Systematic Review and Meta-Analysis'. *Human Reproduction Update*, vol. 27, no. 4, Aug. 2021, pp. 673–96. Silverchair, <https://doi.org/10.1093/humupd/dmab008>.
- Meneghini, Caterina, et al. 'The Impact of Nutritional Therapy in the Management of Overweight/Obese PCOS Patient Candidates for IVF'. *Nutrients*, vol. 15, no. 20, Jan. 2023, p. 4444. www.mdpi.com, <https://doi.org/10.3390/nu15204444>.
- Mizrachi, Yossi, and Dana B. McQueen. 'Embryo Transfer Success: It Is in Our Hands'. *Fertility and Sterility*, vol. 118, no. 5, Nov. 2022, pp. 815–19, <https://doi.org/10.1016/j.fertnstert.2022.08.858>.
- Mohamed Salih, Egbal Lutfi, et al. 'Clinical Predictors of Successful Pregnancy After In Vitro Fertilization (IVF): A Comprehensive Systematic Review of Evidence'. *Cureus*, Jan. 2026. DOI.org (Crossref), <https://doi.org/10.7759/cureus.100734>.
- Penzias, Alan, et al. 'Obesity and Reproduction: A Committee Opinion'. *Fertility and Sterility*, vol. 116, no. 5, Nov. 2021, pp. 1266–85. DOI.org (Crossref), <https://doi.org/10.1016/j.fertnstert.2021.08.018>.
- Petanovski, Z et al. "Impact of body mass index (BMI) and age on the outcome of the IVF process." *Prilozi* vol. 32,1 (2011): 155-71.
- Ribeiro, Silviana, and Mário Sousa. 'In Vitro Fertilisation and Intracytoplasmic Sperm Injection Predictive Factors: A Review of the Effect of Female Age, Ovarian Reserve, Male Age, and Male Factor on IVF/ICSI Treatment Outcomes'. *JBRA Assisted Reproduction*, 2022. DOI.org (Crossref), <https://doi.org/10.5935/1518-0557.20220000>.
- Robin, Fernanda K., et al. 'Donor Oocyte Cycle Characteristics and Outcomes: Factors Potentially Linked with Successful Endings'. *JBRA Assisted Reproduction*, 2022. DOI.org (Crossref), <https://doi.org/10.5935/1518-0557.20220046>.
- Seungdamrong, Aimee, et al. 'Fetal Cardiac Activity at 4 Weeks after in Vitro Fertilization Predicts Successful Completion of the First Trimester of Pregnancy'. *Fertility and Sterility*, vol. 90, no. 5, Nov. 2008, pp. 1711–15. www.fertstert.org, <https://doi.org/10.1016/j.fertnstert.2007.08.018>.
- Shah, Divya K., et al. 'Effect of Obesity on Oocyte and Embryo Quality in Women Undergoing In Vitro Fertilization'. *Obstetrics & Gynecology*, vol. 118, no. 1, July 2011, p. 63. journals.lww.com, <https://doi.org/10.1097/AOG.0b013e31821fd360>.

- Shingshetty, Laxmi, et al. 'Predictors of Success after in Vitro Fertilization'. *Fertility and Sterility*, vol. 121, no. 5, May 2024, pp. 742–51. www.fertstert.org, <https://doi.org/10.1016/j.fertnstert.2024.03.003>.
- Snider, Alexandria P., and Jennifer R. Wood. 'Obesity Induces Ovarian Inflammation and Reduces Oocyte Quality'. *Reproduction*, vol. 158, no. 3, Sept. 2019, pp. R79–90. Silverchair, <https://doi.org/10.1530/REP-18-0583>.
- Yeh, Jason S., et al. 'Pregnancy Rates in Donor Oocyte Cycles Compared to Similar Autologous in Vitro Fertilization Cycles: An Analysis of 26,457 Fresh Cycles from the Society for Assisted Reproductive Technology'. *Fertility and Sterility*, vol. 102, no. 2, Aug. 2014, pp. 399–404. www.fertstert.org, <https://doi.org/10.1016/j.fertnstert.2014.04.027>.

The Effects of Different Cryoprotective Agents (CPAs) on the Cryopreservation of By Aloe Vera Rinds

Cryopreservation allows biological material to be stored at low temperatures, but causes damaging ice crystals to form within preserved material. To combat this, cryoprotectant agents (CPAs) are used. This experiment used a combination of different CPAs, dimethyl sulfoxide (DMSO), glycerol, and n-substituted glycine (NSG/glycine betaine), to determine if their combined use is more viable for the cryopreservation of aloe rinds. The hypothesis, the use of NSG with DMSOs and glycerol would lessen harmful ice crystal formation in aloe, was tested with 9 groups: 8 cryopreserved in different solutions (Distilled water, Glycerol, DMSO, NSG, DMSO+glycerol, DMSO+NSG, Glycerol+NSG, DMSO+glycerol+NSG) and 1 unpreserved group (healthy). With 3 explants per group, 8 groups were stored in an -80°C freezer for one week. The efficacy of the experiment was measured with the explants' regrowth (via regrowth assay), ion leakage (via electrolyte leakage assay), intensity of staining (via TTC assay), and observed cell integrity (via microscopy assay) post-thaw. The electrolyte leakage assay couldn't be used to interpret results due to lack of data collection. This study found that the different combinations of CPAs had no effect on the regrowth ability, metabolic activity, and observed cell integrity, rejecting the hypothesis. These results indicate a need for higher concentrations of CPAs during the cryopreservation process and more specialized assays in order to present more measurable findings that would support what the experiment was set out to study.

Introduction

Cryopreservation is a process in which biological constructs are cooled to very low temperatures in order to preserve them. Cryopreservation techniques are crucial for the long term storage of cells, and have become increasingly helpful. Cell preservation can make important advancements in the field of medicine and agriculture. However, freezing is fatal to most living organisms, as intra- and extracellular ice crystals form, causing mechanical constraints and injury to cells (Jang et al., 2017). In order to combat this issue, the use of cryoprotective agents (CPAs) such as dimethyl sulfoxide (DMSO), glycerol, and N-substituted glycine (NSG) has become more common, though their combined effects have been understudied. DMSO is a cell penetrating cryoprotectant that is most commonly used to preserve animal cells, protecting them from the intracellular formation of ice crystals (Wohnhaas et al., 2019). Glycerol is another popular cryoprotectant, as it also helps combat the mechanical damage caused by ice crystals (Bhattacharya, 2018). NSG is a specific type of peptidomimetic oligomer that is normally synthesized for the purpose of cryopreservation. Aloe vera is an extremely useful medicinal plant with antioxidant and antibacterial properties, helping to reduce dental plaque, accelerate wound healing, and improve skin barrier function (Leech, 2025). Aloe leaf rinds in particular are popularly used as natural insecticides and laxatives to help with constipation (American Chemical Society, 2023). Recently, there have been increased efforts in the cryopreservation of aloe species due to their poor regenerative abilities and the destruction of their natural habitats,

which has limited aloe supply (Berhe et al., 2023). By testing the combined effects of certain cryoprotectants on the viability of cryopreservation, the study of cryopreservation can be improved allowing for the advancement of science and heightened security of all types of plants. Specifically, if using DMSO, glycerol, and NSG together is seen to enhance the viability of cryopreservation and transplantation in aloe, it could help preserve the plant species allowing for its medicinal qualities to be utilizable and more accessible to various fields of study. Therefore, the purpose of this study is to find a unique mix of different CPAs that could enhance the viability of cryopreservation in aloe rinds, and can be used in future cryopreservation of other plants as well. The hypothesis is that the use of NSG with DMSOs and glycerol would lessen harmful ice crystal formation in aloe.

Materials and methods

Materials

- 2mL cryovials
- 1% 2,3,5-triphenyltetrazolium chloride (TTC) (1 g 100% TTC powder in 100 mL distilled water)
- Distilled water
- Deionized water
- -80°C freezer
- 4°C refrigerator
- Scalpel
- Paper towels
- 100x Microscope
- 100% DMSO (using 5 mL with 95 mL of distilled water to get 5%)
- 100% glycerol (using 5 mL with 95 mL of distilled water to get 5%)
- 5M N-substituted glycine (glycine betaine) (diluted with 99.95 mL distilled water and 0.05 mL of 5M n-substituted glycine to get 2.5 mM)
- 6 Mr. Frosty Freezing Containers (18 tube storage each)
- 92% isopropanol alcohol (IPA)
- Microscope slides
- Micropipettes (0-250 μ L and 100-1000 μ L)
- Tweezers
- Permanent markers and tape
- Hotplate
- Ziploc bags
- Conductivity meter
- Conductivity meter solution

Methods

Part A: PPE

1. Wear a lab coat, goggles, and gloves in the lab at all times. Work in the fume hood when needing to work with CPAs. Lastly, use cryo-gloves when handling frozen samples.

Part B: CPA solution prep

1. Prepare the CPA solutions(should be prepared under fume hood):
 1. DMSO/Glycerol solution prep: 5 mL 100% DMSO/Glycerol + 95 mL distilled water, mixed gently.
 1. Add 95 mL of distilled water into a sterile container
 2. Add 5 mL of 100% DMSO/Glycerol into water while stirring gently (under fumehood)
 3. Label it "5% DMSO/Glycerol, date, AS"
 2. Glycerol+DMSO solution; 5 mL 100% DMSO + 5 mL 100% glycerol + 90 mL distilled water, mixed gently
 1. Add 90 mL of distilled water into a sterile container
 2. Add in 5 mL of glycerol
 3. Add in 5 mL of DMSO and mix
 4. Label it "5% Glycerol + 5% DMSO, date, AS"
 3. NSG(glycine betaine) solution: 0.05 mL 5M NSG + 99.95 mL of distilled water
 1. Add 99.95 mL of distilled water into a sterile container
 2. Add 0.05 mL of 5M NSG and mix
 3. Label it "2.5 mM NSG, date, AS"
 4. DMSO + NSG/Glycerol+NSG solution: 0.05 mL 5M NSG + 5.0 mL 100% DMSO/Glycerol + 94.95 mL distilled water
 1. Measure 94.95 mL distilled water into a sterile container
 2. Add 5.0 mL of 100% DMSO/Glycerol slowly while stirring gently (under fume hood)
 3. Add 0.05 mL of 5M NSG (should dilute to 2.5 mM) and mix
 4. Label it "5% DMSO + 2.5 mM NSG/5% glycerol+2.5 mM NSG, date, AS"
 5. Glycerol + DMSO + NSG: 5.0 mL 100% glycerol + 5.0 mL 100% DMSO + 0.05 mL 5 M NSG + 89.95 mL distilled water
 1. Measure 89.95 mL of distilled water into a sterile container
 2. Add 5.0 mL of DMSO slowly while stirring
 3. Add 5.0 mL of glycerol while stirring
 4. Add 0.05 mL of 5M NSG and mix
 5. Label it "5% DMSO + 5% glycerol + 2.5 mM NSG, date, AS"

Part C: Aloe tissue sample preparation

1. Harvest Aloe vera leaves and rinse under tap water, then pat dry.
2. Remove the edge spines of the leaf with a sterile knife

Part D: Reducing aloin in the aloe leaf rinds

1. Put the freshly cut leaves cut end down in a beaker for 20 minutes so the yellow latex can drain from the cut surfaces. Blot the drained latex.
2. Scrape the thin yellow latex-rich layer under the green rind using a sterile scalpel to remove the tissues containing high aloin concentration.
3. Rinse leaf pieces in sterile deionized water to remove surface residue and blot dry with a sterile gauze.

Part E: Explant preparation

1. Cut aloe rind into 5x5 mm explants and label each explant with “reduced aloin”
2. Keep the explants on ice while preparing CPA solutions

Part F: Treatment

1. Place aloe explants into 2 mL cryovials with 1 mL of CPA/NSG/CPA+NSG solution
2. Immediately proceed to freezing.

Part G: Controlled-rate freezing and storage

1. Put the tubes into a Mr. Frosty container
2. Add 92% IPA to the fill line
3. Put the Mr. Frosty containers into a -80°C freezer for a week (the Mr. Frosty will automatically lower the temperature at a constant rate of 1°C/min until it reaches -80°C, allowing the containers to be left in the freezer for 7 days after reaching -80°C)

Part H: Thawing and post-thaw handling

1. Thaw rapidly in a 37°C water bath until the ice melts.
2. Remove the explants and immediately rinse with distilled water to remove residual CPAs and blot dry.
3. Proceed to assays

Assays

TTC Viability Assay (Del Egidio et al., 2017)

- Reads metabolic activity, indicated by a red stain
 - Create a 1% TTC solution with 1 gram of TTC powder and 100 mL of distilled water
 - Rinse thawed explants and place into cryovials containing TTC solutions, with proper labels on vials
 - Place cryovials in 37°C incubator for 36 hours
 - Score intensity of red staining from 0-5 (0 being no stain, 5 being strong red stain)

Electrolyte Leakage assay (Tovar-Rosales et al., 2024) (The data for this assay was unable to be collected in full, so it is not accounted for in results and discussion)

- Assesses membrane integrity by measuring ion leakage into water
 - Rinse thawed explants and place them in deionized water for 30 minutes
 - Take the explants out after 30 minutes, rinse with distilled water, then resubmerge in deionized water for 4 hours
 - (Conductivity measuring steps) Measure the conductivity of each explant group after 4 hours
 - Rinse probe of conductivity meter in distilled water for 30 minutes before measuring
 - Add 4cm of conductivity solution to 2 clean containers (A and B)
 - Swirl the probe in container A and let it stand in container B until the screen of the meter displays “ready”
 - Place the probe in the beaker with the explants whose conductivity are going to be measured
 - When switching between explant groups, (before measuring conductivity) rinse the probe in distilled water for a couple of seconds and follow the conductivity solution steps mentioned above
 - Repeat the conductivity measuring steps for all groups of aloe explants
 - Boil each group of aloe explants for 10 minutes and allow them to cool to room temperature
 - Measure conductivity of each group (refer to conductivity measuring steps)

Regrowth assay (Berhe et al., 2023)

- Determines if the aloe explants are viable after thawing
 - After thawing aloe explants, place each onto moist paper towels and place them in ziploc bags with the group names labeled on each bag
 - Take pictures of growth everyday, with a ruler in each picture, for 7 days to measure growth over time

Microscopy Scoring Assay (Glaser et al., 2018)

- Visual assessment of cell structure and damage
 - Rinse aloe explants in distilled water and place onto microscope slides
 - Observe under 100x light microscope and take photos of what is displayed on the microscope
 - Score the integrity of the cells from a scale of 0-3 (0 having many lysed and empty cells, 3 having mostly healthy cells) Scoring inspired by a method of numerical scoring used in a study on

Actinic Keratosis in which lesions were scored from a range of 0-15 when viewed under a dermatoscope (Soare et al., 2025).

Results and Discussion

Results

TTC Viability Assay (Del Egido et al., 2017)

This assay measures the viability of different CPA solutions in aloe cryopreservation by measuring the metabolic activity, through the intensity of red coloration, in explants from each type of CPA solution. A numerical scale from 0-5 was used (0 being no stain and 5 being strong red stain) to score each explant, with strong red staining indicating the presence of metabolic activity which decreases with the decline in staining intensity (Table 1).

Score (0-5)	TTC scores of each explant in each CPA solution								
	CPA Solution Name								
	Healthy	Distilled Water	Glycerol	DMSO	Glycine Betaine (NSG)	Glycerol+NSG	Glycerol+DMSO	DMSO+NSG	Glycerol+DMSO+NSG
Explant 1	5	0	0	0	0	0	0	0	0
Explant 2	5	0	0	0	0	0	0	0	0
Explant 3	5	0	0	0	0	0	0	0	0

Table 1. TTC scores (from 0-5) of aloe explants treated with different CPA solutions,, made by student researchers. A score of 0 indicates no metabolic activity while a score of 5 identifies a great amount of metabolic activity. Each explant was incubated for 36 hours in a solution of 1% TTC.

All healthy explants scored a 5, indicating the presence of metabolic activity and a strong red stain. Explants preserved with distilled water and the other 7 CPA solution types scored a 0 each, indicating no metabolic activity and no staining (Table 1). The different CPA treatment types had no effect on preventing cell death or damaging ice crystal formation, as none of the explants from each of the groups showed staining or metabolic activity. Thus, the hypothesis is unsupported by the evidence presented in the table. The use of a Kruskal-Wallis test supports this fact, ($H=7.714 < \chi^2$ value =15.507, df=8).

Microscopy Assay (Soare et al., 2025)

This assay measures the viability of different CPA solutions in aloe cryopreservation by scoring each aloe explant's cell integrity through observation under a microscope. The explants are scored on a numerical scale of 0-3 (0 indicating much browning, lysed cells, and the unhealthiest cell integrity while 3 indicates normal cell activity and healthiest cell integrity) (Table 2). This was inspired by a method of numerical scoring used in a study on Actinic Keratosis in which lesions were scored from a range of 0-15 when viewed under a dermatoscope (Soare et al., 2025).

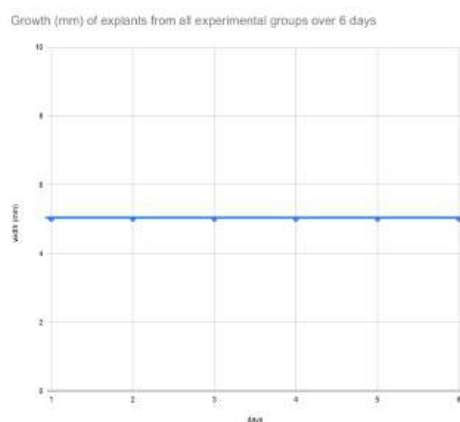
Score (0-3)	Microscopy scoring of each explant in CPA solutions									
	CPA solution name									
	Healthy	Distilled Water	Glycerol	DMSO	NSG	Glycerol+NSG	Glycerol+DMSO	DMSO+NSG	Glycerol+DMSO+NSG	
Explant 1	3	2	1	0	0	1	1	0	2	
Explant 2	2	0	1	0	2	1	0	1	1	
Explant 3	3	2	1	0	0	1	1	0	1	

Table 2, showing the microscopy scores (from 0-3) of each explant from various CPA solution types, made by student researcher. A score of 0 indicates cell lysis and poor cell integrity while a score of 3 indicates healthy cell integrity and no cell death. Each explant was observed under a light microscope and was scored using healthy aloe explants as reference.

The healthy explants showed best cell integrity, with scores ranging from 2-3, while explants from all other groups exhibited unhealthy cell integrity, with scores mainly ranging from 0-2 (Table 2). This indicates that the different CPA solutions didn't protect the cellular integrity of the aloe explants because many aloe explants still showed unhealthy cell composition after being treated with CPAs. This information agrees with the null hypothesis, as it is seen that the use of different CPA solutions has no effect on cell death and harmful ice crystal formation. Doing a Kruskal-Wallis H test helps prove this point once again, ($H=13.35 < \chi^2$ value =15.507, $df=8$).

Regrowth Assay (VivaTowels, 2025)

This assay attempted to measure the viability of different CPA solutions in preventing cell death and the formation of harmful ice crystals during aloe cryopreservation by measuring the growth of explants over the course of 6 days. Explants from each CPA solution group were photographed for 6 days with a ruler in each photo to measure their growth, and graphs were made to keep track of their growth over the days (Figure 1).



Graph 1, showcasing a graph that recorded the growth in mm of explants from all experimental groups over time, made by student researchers. Pictures were taken over a 7 day period, but only 6 days were captured with rulers in each photo to track the growth of the explants. Since there was no difference in growth amongst all experimental groups, one graph was sufficient to record the data for all groups.

The graphs above show how none of the explants from any group, including the healthy explants, grew, which may indicate an error in the protocol (Figure 1). Since the method to

regrow the explants was inspired by seed germination, it may not have been suitable for aloe explants as seeds are thicker and have different structures and properties than plant leaf explants. This suggests that the results may be due to an error in the protocol and aren't valid to properly assess post-thaw growth of explants. Regardless, the information still supports the null hypothesis as the lack of growth across 6 days show how the different CPA solution groups had no effect on preserving the health of aloe explants, as there was no difference in the growth of explants treated with CPAs and explants in the control group (healthy and distilled water).

Discussion

Hypothesis Evaluation: The results from all the assays do not support the hypothesis and instead support the null hypothesis. Throughout the microscopy, TTC, and regrowth assays, the results of the aloe explants that were frozen in the different CPA solutions showed no difference in results compared to the group of explants frozen in distilled water. Both the TTC and regrowth assays showed no difference in the width or metabolic activity of aloe explants that were cryopreserved in distilled water and the 7 different CPA solutions, suggesting that the different combinations of CPAs did not keep the aloe explants alive during the freezing process nor did they contribute to the regrowth ability of the aloe explants after thawing. The microscopy and TTC assays also work together to support the null hypothesis as the calculated H values after doing a Kruskal-Wallis test with the data from both assays reveals that the chi-squared value at 8 degrees of freedom (9 groups total-1) was higher than the calculated H values. This indicates that the null hypothesis is supported while the hypothesis is rejected. As you can see, evidence from all three assays overlap and support the null hypothesis rather than the hypothesis.

Context and Comparison: Cryopreservation allows the long-term storage of biological material, but there are still hundreds of cell types that cannot be cryopreserved (Bryant et al., 2023). The purpose of this experiment was to find a new mix of different CPAs that could increase the viability of cryopreservation in aloe explants and could also be used on new, not yet preserved plants. The specific CPAs chosen for this experiment included glycerol and DMSO for their common use in cryopreservation throughout the years, and glycine betaine as it is a new and promising peptidomimetic oligomer that has similar functions as other successful CPAs. Glycerol is nontoxic and has a high degree of biosafety, and is a promising solution for cryopreservation (Zhang et al., 2022) and DMSO has also been extensively used as a cryoprotectant due to its ability to protect cells during freezing and thawing (strexcell, 2023). However, the results of the microscopy, TTC, and regrowth assays strongly suggest that the different combinations of the three CPAs used in this experiment did nothing to protect the aloe explants' cell integrity, growth, and survivability during cryopreservation. The unsuccessful experiment may have been due to using the CPAs DMSO and glycerol, which are largely used to freeze mammalian cells, in the cryopreservation of plant cells. Numerous studies have reported that exogenous glycine betaine improved plant tolerance against chilling (Min et al., 2021). DMSO is a cell penetrating cryoprotectant that is most commonly used to preserve animal cells,

protecting them from the intracellular formation of ice crystals (Wohnhaas et al., 2019) while glycerol is largely used in the preservation of sperm cells (Lin et al., 2023). This suggests that the field of cryopreservation will continue to use common CPAs such as glycerol, DMSO, and glycine betaine on their own rather than in conjunction, as their solo use has already been proven to be viable for cryopreserving a multitude of cell types.

Error Analysis: The experiment may not have supported the hypothesis due to incorrect concentrations of cryoprotectant agents (CPAs). Aloe vera rinds are 15-20 cells thick, which makes it harder for chemicals to penetrate and become effective, requiring a higher concentration of CPAs to actually have an effect on cryopreservation (Rahman et al., 2017). It was found that a majority of scientists utilize a concentration of 5% up to 10% in their freezing medium (Essani, 2024). Additionally, red blood cells are normally cryopreserved using 20–40% glycerol and immortalized cell lines are routinely stored in 10% DMSO (Gibson and Murray, 2022). Since this experiment only used a 5% concentration of CPAs, the concentration may have been too low, especially for explants that are as thick as aloe leaf rinds, as it seems the normal concentration used of DMSO and glycerol in cryopreservation ranges from 10%-40%. For future experiments related to the topic of this experiment, the use of a 10% concentration or higher for CPA solutions would be recommended and preferred. Additionally, the regrowth assay may have not been successful as the experiment used a protocol similar that allowed the germination of seeds at home. Since seeds and aloe explants are different in terms of structure and function, the protocol for germinating seeds was not compatible for regrowing aloe explants. For this reason, future experiments should find a method of regrowth that is specific for regrowing plant leaf explants in order to be more suitable for aloe explant regrowth.

Conclusion: There were no significant differences in the results of the experiment. The TTC and regrowth assays showed no difference in the width or metabolic activity of aloe explants in different CPA solutions (Table 1 and Graph 1) and the microscopy and TTC assays do not support the hypothesis as the use of a Kruskal Wallis test fails to reject the null hypothesis ($H=7.714 < \chi^2 \text{ value} = 15.507$, $df=8$) (Table 1) ($H=13.35 < \chi^2 \text{ value} = 15.507$, $df=8$) (Table 2). This indicates that there was no difference between the different CPA treatments' abilities in maintaining the viability of aloe explants during and after cryopreservation.

Works Cited

- American Chemical Society. (2023, August 15). Discarded aloe peels could be a sustainable, natural insecticide (video). American Chemical Society.
<https://www.acs.org/pressroom/newsreleases/2023/august/discarded-aloe-peels-could-be-sustainable-natural-insecticide-video.html>
- Berhe, B. D., Sbhatu, D. B., Munawar, T. M., & Gebreyohannes, G. (2023). Aloe monticola Reynolds: A refugee of the mountains - contributing towards its conservation through in vitro propagation. *Heliyon*, 9(12), e22955. <https://doi.org/10.1016/j.heliyon.2023.e22955>
- Bhattacharya, S. (2018). Cryoprotectants and Their Usage in Cryopreservation Process. IntechOpen. doi: 10.5772/intechopen.80477
- Bryant, S. J., Elbourne, A., Greaves, T. L., & Bryant, G. (2023). Applying soft matter techniques to solve challenges in cryopreservation. *Frontiers in Soft Matter*, 3, Article 1219497. <https://doi.org/10.3389/frsfm.2023.1219497>
- Essani, K. (2024, October 11). DMSO: Freezing cells. Retrieved from <https://www.susupport.com/knowledge/biopharmaceutical-products/cell-gene-therapy/dmso-freezing-cells>
- Glaser, A. K., Reder, N. P., Chen, Y., McCarty, E. F., Yin, C., Wei, L., Wang, Y., True, L. D., & Liu, J. T. C. (2017). Light-sheet microscopy for slide-free non-destructive pathology of large clinical specimens. *Nature biomedical engineering*, 1(7), 0084. <https://doi.org/10.1038/s41551-017-0084>
- Jang, T. H., Park, S. C., Yang, J. H., Kim, J. Y., Seok, J. H., Park, U. S., Choi, C. W., Lee, S. R., & Han, J. (2017). Cryopreservation and its clinical applications. *Integrative medicine research*, 6(1), 12–18. <https://doi.org/10.1016/j.imr.2016.12.001>
- Leech, J. (2025, April 16). What are the benefits of aloe vera? Medical News Today. Retrieved September 8, 2025, from <https://www.medicalnewstoday.com/articles/318591>
- Lin, H. H., Mermillod, P., Grasseau, I., Brillard, J. P., Gérard, N., Reynaud, K., Chen, L. R., Blesbois, E., & Carvalho, A. V. (2023). Is glycerol a good cryoprotectant for sperm cells? New exploration of its toxicity using avian models. *Animal reproduction science*, 258, 107330. <https://doi.org/10.1016/j.anireprosci.2023.107330>
- Lopez Del Egidio, L., Navarro-Miró, D., Martinez-Heredia, V., Toorop, P. E., & Iannetta, P. P. M. (2017). A spectrophotometric assay for robust viability testing of seed batches using 2,3,5-triphenyl tetrazolium chloride: Using *Hordeum vulgare* L. as a model. *Frontiers in Plant Science*, 8, Article 747. <https://doi.org/10.3389/fpls.2017.00747>
- Min, K., Cho, Y., Kim, E., Lee, M., & Lee, S.-R. (2021). Exogenous glycine betaine application improves freezing tolerance of cabbage (*Brassica oleracea* L.) leaves. *Plants*, 10(12), 2821. <https://doi.org/10.3390/plants10122821>
- Murray, K. A., & Gibson, M. I. (2022). Chemical approaches to cryopreservation. *Nature Reviews Chemistry*, 6(8), 579–593. <https://doi.org/10.1038/s41570-022-00407-4>

- Rahman, S., Carter, P., & Bhattarai, N. (2017). Aloe Vera for Tissue Engineering Applications. *Journal of Functional Biomaterials*, 8(1), 6. <https://doi.org/10.3390/jfb8010006>
- Soare, C., Cozma, E. C., Giurcăneanu, C., & Voiculescu, V. M. (2025). Comparative Analysis of Clinical, Dermoscopic, and Confocal Microscopy Scores for Assessing Severity of Actinic Keratosis. *Cancers*, 17(17), 2899. <https://doi.org/10.3390/cancers17172899>
- Strex Cell. (2023, May 5). Freezing cells in DMSO: Benefits & process explained. Retrieved from <https://strexcell.com/blog/freezing-cells-in-dms0-the-benefits-and-process-explained/>
- Tovar-Rosales, L., Arthikala, M. K., & Nanjareddy, K. (2024). Electrolyte leakage assay to analyze membrane integrity in leaves (v1) [Protocol]. *protocols.io*. <https://doi.org/10.17504/protocols.io.ewov19j37lr2/v1>
- Viva® Towels. (n.d.). How to germinate seeds in paper towels. Retrieved November 28, 2025, from <https://www.vivatowels.com/en-us/cleaning-tips/everyday-cleaning/germinating-seeds-paper-towel>
- Wohnhaas, C.T., Lepar, G.G., Fernandez-Albert, F. et al. DMSO cryopreservation is the method of choice to preserve cells for droplet-based single-cell RNA sequencing. *Sci Rep* 9, 10699 (2019). <https://doi.org/10.1038/s41598-019-46932-z>
- Zhang, P.-Q., Tan, P.-C., Gao, Y.-M., Zhang, X.-J., Xie, Y., Zheng, D.-N., Zhou, S.-B., & Li, Q.-F. (2022). The effect of glycerol as a cryoprotective agent in the cryopreservation of adipose tissue. *Stem Cell Research & Therapy*, 13, Article 152. <https://doi.org/10.1186/s13287-022-02817-z>

Foreign Exchange Intervention and UAH/USD Volatility in Ukraine: Evidence Before and During the 2022-2026 War By Rynat Yushkov

Abstract

Exchange rate volatility is a major concern for emerging economies, especially during crises. Following Russia's full-scale invasion of Ukraine in February 2022, the National Bank of Ukraine (NBU) relied heavily on foreign exchange (FX) interventions while also shifting the exchange rate regime. Using daily official UAH/USD data (01 Aug 2018–15 Jan 2026), this paper computes log returns and measures volatility via (i) regime-specific standard deviations and (ii) a 30-observation rolling standard deviation averaged to a monthly frequency to align with the NBU's monthly net FX interventions (purchases minus sales). Volatility differs sharply by regime: it is highest pre-war (0.0045), mechanically near zero under the two pegs (0.0005 and 0.0003), and increases under managed flexibility (0.0026). Intervention volumes rise markedly after February 2022, with net FX sales as the predominant component. Monthly intervention–volatility co-movement is positive, consistent with interventions responding to stress rather than providing clear causal evidence of stabilisation. A simple post-month check for the ten largest net-sale months shows volatility falls in 6 cases but rises in 4, providing mixed support for short-run dampening and limited evidence that effectiveness diminishes over time. The analysis remains descriptive given endogeneity and the monthly reporting of interventions.

1. Introduction

Exchange rate volatility is widely regarded as a significant source of macroeconomic risk, particularly for emerging economies. Sudden and significant movements in exchange rates can make it easier for inflationary pressures to build and increase uncertainty for firms and households. This can make it difficult for monetary policy to be effective. As a result, central banks use foreign exchange interventions as an instrument to reduce high exchange rate variability. The effectiveness of such interventions, however, remains inconclusive in the literature, particularly in periods of high uncertainty.

Ukraine offers a particularly relevant case for this debate. When the war began in February 2022, the economy was hit by a series of sudden shocks. Trade was disrupted, capital began to flow out of the country, infrastructure was damaged, and investor confidence weakened sharply. Together, these developments put strong pressure on the Ukrainian hryvnia and led to instability in the exchange rate. In response, the National Bank of Ukraine (NBU) implemented a broad set of monetary measures aimed at maintaining financial stability, including a sharp increase in the key policy rate, liquidity support for banks, foreign exchange interventions, and temporary currency restrictions to protect international reserves and limit panic-driven depreciation.¹

Prior research suggests that, in the short-run, exchange rates are not driven solely by economic fundamentals. Market behaviour, different levels of information, as well as differing exchange rate expectations among market participants, can result in exchange rate deviations

from economic fundamentals.² Under such circumstances, the influence of a central bank on exchange rate movements depends not only on direct market interventions but also on communication mechanisms aimed at shaping market expectations. High levels of uncertainty can lead to negative outcomes due to poorly anchored market expectations; conversely, clear policy guidelines can help moderate self-reinforcing market behaviour.²

However, one important empirical issue that makes assessments of FX interventions' effectiveness less clear is related to endogeneity. FX interventions are not independent events, unrelated to market conditions. Rather, central banks usually react to exchange rate pressure, rising volatility, or heightened uncertainty, meaning interventions are more likely to be observed precisely in periods of market stress. Additionally, there is a degree of "circularity" in analysing interventions and volatility; that is, volatility may change following interventions, but at the same time, intervention volumes tend to increase when volatility rises. Therefore, a positive association between interventions and volatility would not necessarily reflect a direct effect of interventions on volatility, but could instead capture the NBU's response to market pressures, which complicates causal interpretation.³

Moreover, the war caused a significant change in Ukraine's monetary policy framework. Before 2022, the NBU largely operated under an inflation-control regime based on interest rates and expectations. In the first months of the war, however, these market-based instruments lost some of their effectiveness due to psychological shocks, high liquidity, and uneven liquidity distribution across banks. As a result, tools such as foreign exchange interventions, capital controls, and exchange rate fixing became the primary means of maintaining macro-financial stability, while the policy rate played a more complementary role.⁴ Although short-term stability was restored, its impact was uneven and often delayed, reflecting significant time lags.⁴

Even though the NBU relied heavily on foreign exchange interventions during the war, evidence on their impact on exchange rate volatility remains very limited, especially in a setting where intervention volumes increased sharply, and the exchange rate regime changed. While some researchers indicate that intervention and communication policies can stabilise markets during periods of turmoil or crisis, it is still unclear whether such policies effectively reduce uncertainty when shocks are persistent, as during wartime. It follows that wartime conditions may alter the effectiveness of standard policy tools, which makes examining the NBU's intervention, volatility relationship during this period particularly relevant.

In this paper, the focus is on how the foreign exchange intervention policy of the National Bank of Ukraine has influenced the volatility of the UAH/USD exchange rate. A comparison between exchange rates before and during the Russia–Ukraine war will help determine whether there is a relationship between foreign exchange intervention policy and the level of exchange rate volatility, and whether this relationship changes from the pre-war period to the wartime period. It is hypothesised that foreign exchange interventions were successful in reducing exchange rate volatility in the short-run, although their impact gradually declined during the war period.

To operationalize the analysis, the sample is divided into four exchange rate regimes: the pre-war period, Peg 1 (introduced in February 2022), Peg 2 (after the July 2022 adjustment), and the subsequent phase of managed flexibility (from October 2023 onward). Volatility is measured in two ways: (i) regime-specific standard deviations of daily returns, and (ii) a rolling standard deviation of returns (computed over a fixed window) that traces how volatility evolves over time and around major intervention episodes.

2. Materials and Methods

2.1 Data and sample period

In order to assess whether foreign exchange interventions affected exchange rate volatility during the war, it is necessary to shift the focus from the theoretical discussion to an empirical examination of the data. This paper uses the official UAH/USD exchange rate at daily frequency over the period 2018 to 2026.⁵ Foreign exchange intervention data are taken from the National Bank of Ukraine (NBU) and are used at a monthly frequency consistent with how intervention statistics are reported.⁶

2.2 Exchange rate regimes

Because the exchange rate was governed by different policy regimes across the sample rather than one consistent market-determined process, the analysis distinguishes between four main settings: the pre-war period (01 Aug 2018–23 Feb 2022), the first fixed-rate peg introduced after the start of the war (Peg 1; 24 Feb 2022–20 Jul 2022), the second fixed-rate peg introduced after the July 2022 adjustment (Peg 2; 21 Jul 2022–02 Oct 2023), and the subsequent managed-flexibility regime (from 03 Oct 2023 onward).^{7,8,9} These regimes were used for regime-level volatility comparisons and for interpreting movements in volatility over time.

Table 1. Exchange rate regimes and sample sizes (daily observations).

Regime	Period	Policy setting	Daily observations
Pre-war	01 Aug 2018–23 Feb 2022	Managed / inflation-control framework	883
Peg 1	24 Feb 2022–20 Jul 2022	Fixed peg (~29.25 UAH/USD)	147
Peg 2	21 Jul 2022–02 Oct 2023	Fixed peg (36.5686 UAH/USD)	439
Managed flexibility	03 Oct 2023–15 Jan 2026	Managed flexibility	598

Table 1 summarises the regime splits used in the analysis and reports the sample size (number of daily official-rate observations) in each regime. These counts indicate how much data underlie each regime-based volatility estimate.

2.3 Exchange rate returns

In order to move beyond a descriptive analysis of exchange rate levels and toward a systematic examination of exchange rate instability, the analysis focuses on exchange rate returns

and their volatility. The exchange rate returns are calculated through logarithmic differences of the UAH/USD exchange rate: $r_t = \ln(P_t) - \ln(P_{t-1})$

In this equation, P_t refers to the official UAH/USD exchange rate on day t , while P_{t-1} is the exchange rate from the previous trading day. Logarithmic returns are commonly used when working with exchange rate data because they make changes easier to compare over time. Using log returns also helps deal with large swings in the data, which are more likely to appear during periods of high volatility.

2.4 Volatility measures

Exchange rate volatility is calculated in two ways.

(1) Regime-based volatility (Figure 2). Volatility is calculated as the standard deviation of daily log returns within each regime:

$$\sigma_k = \sqrt{(\sum_{i=1}^{n_k} (r_i - \mu_k)^2) / (n_k - 1)}$$

where k denotes the exchange rate regime, r_i is the daily log return in regime k , μ_k is the mean return within that regime, and n_k is the number of daily observations in regime k .

(2) Time-varying volatility (Figure 3). To capture changes in volatility over time, this study applies a 30-observation rolling standard deviation to daily returns:

$$\sigma_t = \sqrt{(1/29) \cdot \sum_{i=0}^{29} (r_{t-i} - \mu_t)^2)}$$

where σ_t is the 30-observation rolling standard deviation of returns on day t . The sum runs over the most recent 30-observation returns $r_t, r_{t-1}, \dots, r_{t-29}$, each measured relative to the 30-observation mean μ_t . The squared deviations are averaged using $1/29$, and the square root gives the standard deviation. The rolling window is defined over the most recent 30 days in the available daily series. During the peg periods, the official rate is reported on calendar days, so the rolling window corresponds to 30 calendar days. Outside the peg periods, the official series is primarily reported on trading days, so the same 30-observation window corresponds to approximately 30 trading days.

2.5 NBU foreign exchange intervention measure

With exchange rate volatility measured, the main policy variable in this study is the NBU's foreign exchange intervention. Net intervention is measured monthly as:

$$Net\ Intervention_m = FX\ Purchases_m - FX\ Sales_m$$

Negative values indicate months in which the NBU was a net seller of foreign currency, while positive values indicate net purchases and reserve accumulation.

2.6 Aligning monthly interventions with daily volatility

Because interventions are observed monthly while volatility is constructed from daily exchange rate returns, the two series must be aligned before they can be compared. To do this, the rolling volatility measure is summarised at the monthly level by taking the average of daily rolling-volatility observations within each month. This creates a monthly volatility series that can

be plotted alongside monthly net FX intervention. Using this monthly alignment, “diminishing effectiveness over time” is interpreted as a weaker tendency for volatility to fall in the month following large net FX sales in later war years compared with earlier periods.

2.7 Limitations and data constraints

Because volatility is computed from the official exchange rate, volatility during peg periods may be mechanically low due to policy constraints on day-to-day exchange rate movement. In addition, simple correlations between net interventions and volatility may be vulnerable to reverse causality, since the NBU is more likely to intervene in months when market pressure and volatility are already elevated. In addition, intervention data are reported monthly, so comparisons of volatility must be interpreted as broader co-movements rather than one-to-one daily relationships. Finally, wartime conditions can create gaps between the official and market exchange rates (e.g., cash or parallel-market rates). Using only the official rate may therefore understate the volatility actually experienced by households and firms, especially during periods of tight controls or reduced market liquidity.

3. Results

This section proceeds in four steps. First, Figure 1 establishes the major exchange rate regime shifts across the sample. Second, Figures 2 and 3 measure volatility using regime-based and rolling standard deviations to show how instability differs across regimes and over time. Third, Figure 4 documents the scale and direction of NBU net FX interventions. Finally, Figure 5 aligns monthly interventions with monthly-averaged rolling volatility, and Table 2 provides a simple post-month check for whether volatility tends to fall after the largest net FX sales.

3.1 Exchange rate regimes and level dynamics

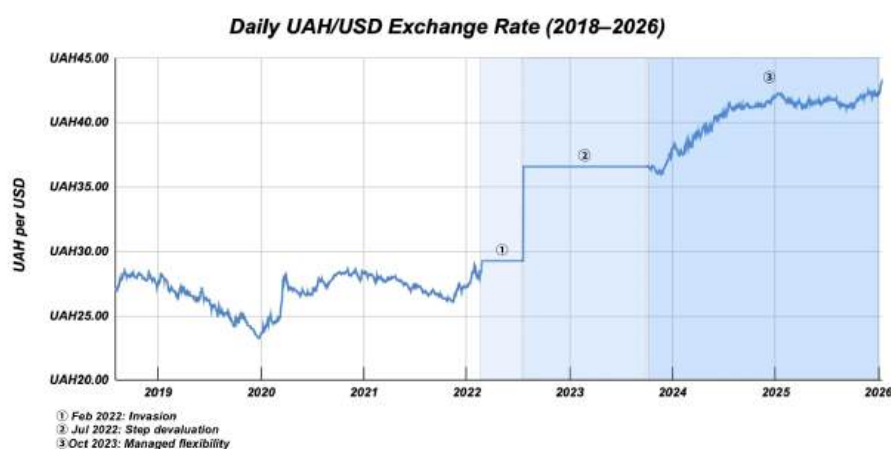


Figure 1: Official UAH/USD exchange rate, Aug 2018 – Jan 2026 (daily).⁵

Figure 1 depicts the official UAH/USD exchange rate from 2018 to 2026. Until the full-scale invasion, the exchange rate was relatively stable, staying below 29 UAH/USD for most

of the period. Following the outbreak of the war in February 2022, the NBU adopted a fixed exchange rate regime, setting the rate at around 29.25 UAH/USD to reduce uncertainty and slow depreciation pressures.⁷ Later, in July 2022, the NBU implemented a discrete one-off adjustment, devaluing the official rate to 36.5686 UAH/USD.⁸ This policy change appears in Figure 1 as a clear vertical step upward, after which the exchange rate remained essentially flat at the new peg until October 2023. Since October 2023, it can be noted that the NBU shifted into a managed flexibility regime, where exchange rate adjustments resumed, and the hryvnia began to depreciate more gradually with more frequent day-to-day variation.⁹

This figure provides the necessary background for the volatility analysis that follows, because it shows that the exchange rate was governed by different policy regimes across the sample rather than one consistent market-determined process. As a result, the analysis moves beyond the level of the exchange rate and instead focuses on returns and volatility measures, which allow short-run instability to be compared more systematically across periods. For the remainder of the paper, the February 2022 shift to a fixed exchange rate, the July 2022 discrete devaluation, and the October 2023 shift to managed flexibility are treated as key policy episodes, since they plausibly change both the behaviour of the official exchange rate and the conditions under which interventions operate.

3.2 Volatility across regimes

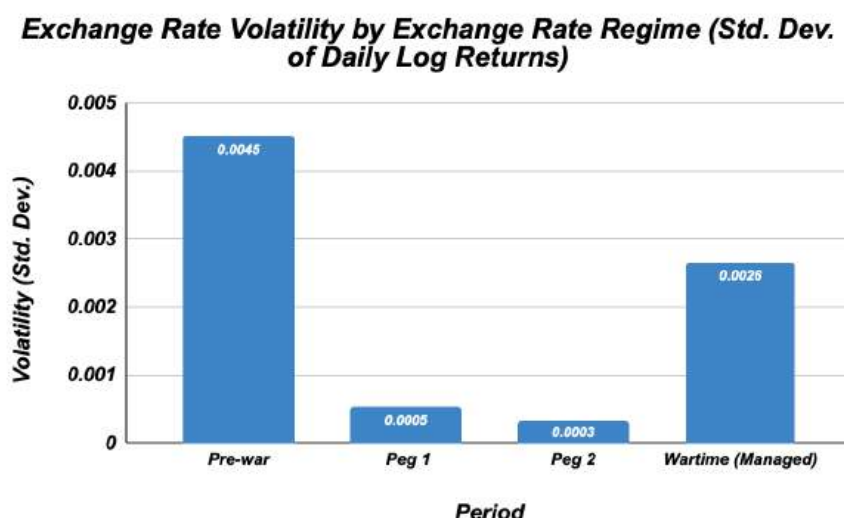


Figure 2: Standard deviation of daily log returns by exchange rate regime.

In order to move beyond the descriptive movement of the exchange rate and toward a systematic examination of exchange rate instability, volatility is calculated from daily logarithmic returns. Figure 2 reports the resulting volatility estimates across the main exchange rate regimes examined in this paper: the pre-war period, Peg 1, Peg 2, and the

managed-flexibility regime. The pre-war period (01 Aug 2018–23 Feb 2022) displays the highest volatility (0.0045).

Lower volatility during the fixed exchange rate periods (Peg 1 and Peg 2) is expected because day-to-day movements in the official exchange rate are mechanically constrained under a peg. Consistent with this, volatility falls to 0.0005 during Peg 1 (24 Feb 2022–20 Jul 2022) and remains close to zero at 0.0003 during Peg 2 (21 Jul 2022–02 Oct 2023). Volatility is higher under the managed-flexibility regime, reflecting the resumption of more frequent exchange rate adjustments; from 03 Oct 2023 onward, volatility rises again to 0.0026. At this stage, the figure provides a regime-based benchmark for exchange rate instability, which is necessary before examining changes in volatility over time and introducing foreign exchange interventions.

3.3 Rolling volatility over time

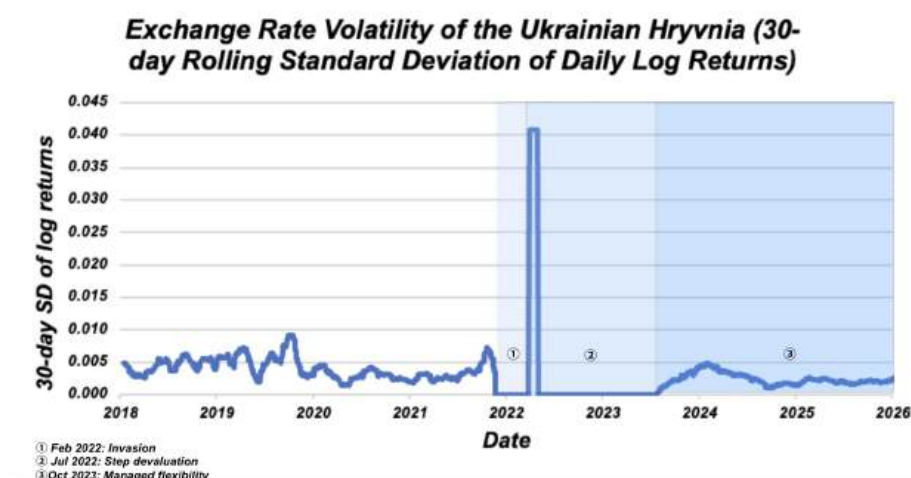


Figure 3: 30-observation rolling standard deviation of daily log returns (UAH/USD).

While Figure 2 is useful for comparing the degree of exchange rate volatility across regimes, it does not depict changes in volatility over time or how quickly volatility adjusts to shifting economic conditions. To address this limitation, a more dynamic measure is constructed by applying a 30-observation rolling standard deviation to daily exchange rate returns, which captures short-run movements in exchange rate variability. Figure 3 plots this 30-observation rolling volatility measure for the UAH/USD exchange rate over the sample period.

Figure 3 shows that the 30-observation rolling volatility remained relatively low throughout most of the pre-war period, indicating that day-to-day exchange rate returns were generally contained. Around the start of the full-scale invasion in 2022 (February 24th), volatility changed sharply: once the official exchange rate was pegged, daily movements in the rate became minimal, causing rolling volatility to fall close to zero. A pronounced spike then appears when the exchange rate experiences a discrete adjustment, since the rolling 30-observation window temporarily includes mostly near-zero returns alongside a single large return, increasing dispersion in the series. As the window moves forward, this one-off jump drops out of the

30-observation period, and volatility declines again. In the latter part of the sample, volatility remains above zero but relatively moderate compared with the July 2022 spike and relative to the higher volatility observed in parts of the pre-war period, consistent with a return to limited day-to-day exchange rate variation under a more flexible regime.

3.4 NBU foreign exchange interventions over time

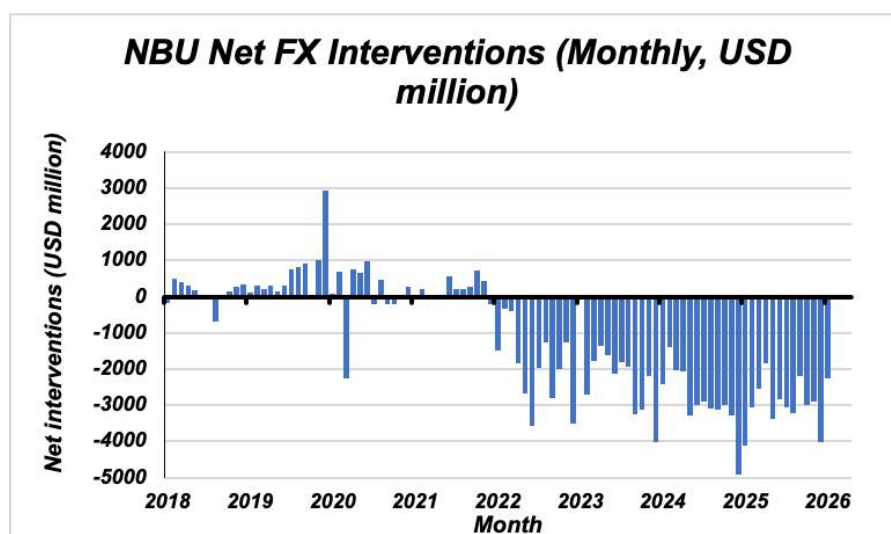


Figure 4: Monthly NBU net FX interventions (USD million), 2018–2026.⁶

With exchange rate volatility measured over time in the previous section, the analysis turns to the main policy variable in this study: foreign exchange intervention by the National Bank of Ukraine (NBU). Figure 4 presents the NBU's monthly net FX intervention series (USD million) from 2018 to 2026. The amounts represent the intervention implemented as measured by FX purchases minus FX sales for each month. A monthly frequency is used because this is how intervention data are reported for Ukraine, making it meaningful for comparison across the sample.

Negative net intervention indicates months in which the NBU was a net seller of foreign currency, whereas positive values indicate net purchases and reserve accumulation. Before the full-scale invasion, interventions were relatively smaller and mixed in sign, with both net purchases and net sales across months. During the wartime period, interventions become persistently negative, with several large net-sale months (visible as deep troughs), indicating sustained FX sales to support the currency and manage reserves under heightened pressure. The figure provides the intervention context needed before comparing intervention activity with changes in exchange rate volatility.

3.5 Intervention activity and monthly exchange rate volatility

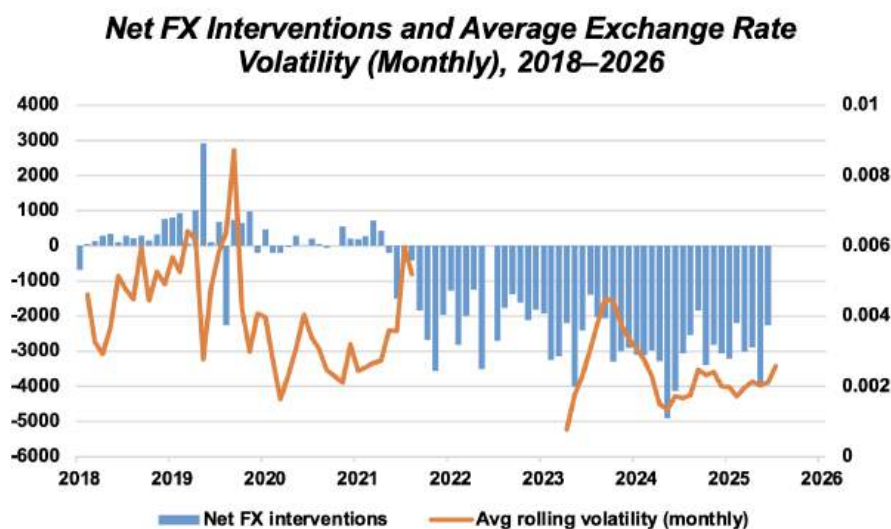


Figure 5: Monthly net FX interventions (USD million) and average monthly rolling volatility.⁶

Because interventions are recorded monthly while volatility is calculated daily, the two series must be aligned before they can be compared. To do this, the rolling volatility measure from Figure 3 is summarised at the monthly level by taking the average of daily rolling-volatility observations within each month. Figure 5 plots monthly net FX intervention alongside average monthly exchange rate volatility. In the data, the two series show broad co-movement: months with larger net FX sales (more negative intervention values) tend to coincide with higher volatility, while periods of smaller sales or reduced intervention intensity generally align with lower volatility, consistent with interventions responding to episodes of market stress. Figure 5 is best interpreted as descriptive evidence of regime-dependent co-movement rather than visual proof that larger net sales coincide with higher volatility.

This comparison provides an initial descriptive basis for assessing how intervention intensity is associated with volatility. The positive co-movement observed is consistent with interventions responding to market stress, rather than providing clear visual evidence of a stabilisation effect. At the same time, because the intervention series is monthly and the volatility series is derived from daily returns, the relationship is interpreted in terms of broader co-movements around the same periods rather than a one-to-one daily match.

3.6 A simple check of short-run volatility after large intervention months

Figure 5 shows broad co-movement between net FX interventions and volatility, but this does not directly test the hypothesis. To make the evidence more systematic, the analysis identifies the largest net-sale months (months with the most negative net FX intervention values) and compares average monthly rolling volatility in the intervention month with average monthly rolling volatility in the following month. If interventions reduce volatility in the short run, volatility should fall in the month after large net FX sales. Table 2 reports this comparison for the largest net-sale months and indicates whether volatility declines in the post-month.

Table 2. Average monthly rolling volatility in the intervention month vs the following month (largest net-sale months).

Month (YYYY-MM)	Net FX intervention (USD mn)	Avg monthly rolling volatility (month)	Avg monthly rolling volatility (next month)	Change (next – month)
2024-12	-4904.32	0.001336	0.001722	0.000387
2025-01	-4129.33	0.001722	0.001668	-0.000054
2025-12	-4024.75	0.002021	0.002122	0.000101
2023-12	-4013.49	0.001762	0.00231	0.000548
2025-05	-3392.26	0.002314	0.002416	0.000102
2024-05	-3296.25	0.004411	0.003766	-0.000645
2024-11	-3282.38	0.001501	0.001336	-0.000166
2025-08	-3213.54	0.001982	0.001725	-0.000257
2024-09	-3110.65	0.002761	0.002283	-0.000478
2024-08	-3097.74	0.003066	0.002761	-0.000305

Table 2 shows that volatility declines in 6 of the 10 largest net-sale months (negative values in the Change column), which is consistent with short-run stabilisation in some episodes. However, volatility increases in the remaining four cases, indicating that the post-intervention decline is not a consistent pattern. This mixed result provides only partial support for the hypothesis and is consistent with the descriptive nature of the analysis: large intervention months may coincide with shifting shocks and expectations, and monthly data do not isolate the timing of intervention effects. A weaker tendency for volatility to fall in the post-month in later war years would be consistent with diminishing effectiveness over time. This post-month comparison is intended as a structured check rather than an identification strategy; stronger causal inference would require higher-frequency intervention timing or an event-study/regression framework.

3.7 Results summary and link back to the hypothesis

Overall, the results show that exchange rate volatility in Ukraine differs strongly across policy regimes. The regime-based volatility values (Figure 2) and the rolling volatility series (Figure 3) both indicate that volatility is mechanically low during fixed exchange rate periods, while it becomes more noticeable once the NBU returns to a more flexible exchange rate framework. The rolling measure also highlights that volatility responds sharply to discrete exchange rate adjustments, producing temporary spikes even when the exchange rate is otherwise constrained.

When the intervention series is compared with the monthly volatility measure (Figure 5) and then checked more systematically (Table 2), the evidence provides partial support for the hypothesis: volatility falls in some large net-sale episodes, but not consistently. The patterns are consistent with the idea that interventions can help stabilize markets in the short-run, particularly

around periods of stress, but the relationship becomes less clear as wartime conditions persist and the exchange rate framework changes. These results, therefore, motivate the Discussion section, where the findings are interpreted relative to the hypothesis and linked to the limits of policy effectiveness under prolonged uncertainty.

3.8 Interpreting the findings relative to the hypothesis

This paper hypothesised that NBU foreign exchange interventions reduced hryvnia volatility in the short-run, but that their effectiveness weakened as the war continued. Based on the figures, this hypothesis is partially supported. Figures 2 and 3 show much lower volatility during the fixed exchange rate periods, while volatility increases once the NBU moves toward a more flexible framework. This aligns with the subsequent policy shift to managed exchange rate flexibility (effective 3 October 2023), which allowed more frequent movements in the exchange rate while still keeping the NBU actively involved in stabilisation.⁹ At the same time, the peg itself limits daily changes in the official rate, so the low volatility in Peg 1 and Peg 2 should not be taken as evidence of an intervention effect by itself. Instead, it reflects a wider wartime setup, where the fixed exchange rate and restrictions were doing most of the stabilising, and interventions were used alongside them.^{10,11} Because the peg mechanically constrains movements in the official rate, intervention effectiveness is difficult to isolate from peg-period volatility alone and must be inferred primarily from the managed-flexibility phases, when the exchange rate is allowed to adjust, and interventions can be assessed against observable fluctuations.

3.9 Exchange rate regime mechanics and the discrete adjustment

The rolling volatility plotted in Figure 3 drops close to zero after the peg is introduced, which is expected because the official rate is no longer moving day by day in the same way. The spike around July 2022 is also consistent with what happened in policy terms: the NBU carried out a one-time devaluation and fixed the official rate at 36.5686 UAH/USD.⁸ Because volatility is calculated using a rolling window, one large one-day change within a window of near-zero returns temporarily pushes the rolling standard deviation up. Once that one-off change leaves the 30-observation window, the volatility measure falls back again.

4. Discussion

The NBU net intervention data show strong month-to-month variation across the sample.⁶ When this is compared with monthly average rolling volatility (Figure 5), the series exhibits broad co-movement: larger net FX sales tend to coincide with higher volatility, consistent with interventions responding to market stress rather than providing clear visual evidence of a stabilisation effect.^{9,12} In other words, because official-rate volatility was mechanically compressed under the peg, it was difficult to disentangle the marginal effect of interventions from the regime itself, and this became even harder later in the war as repeated shocks to trade flows, capital movements, and expectations continued to reshape market conditions.

This also matches expectation-based views in the literature. Under high uncertainty, exchange rates can deviate from fundamentals because beliefs, information gaps, and central-bank communication become more important for market behaviour.² That makes it plausible that interventions may still help at certain points, but the stabilising relationship became less consistently observable over time as expectations keep shifting and the operating policy regime changes. A further reason the stabilising association may weaken over time is that prolonged net FX sales gradually tighten reserve buffers, reducing the NBU's capacity to intervene at the same scale without raising concerns about sustainability. As the shock becomes persistent rather than temporary, interventions are also more likely to be interpreted as signalling stress, which can shift expectations and blunt the calming effect. Credibility matters as well: if market participants expect future devaluations or regime adjustments, short-run smoothing is harder to maintain. Finally, as capital controls are adjusted over time, changes in access to FX and market depth can alter how interventions transmit into observed volatility, making the relationship less stable across regimes.

4.1 Policy takeaway

Taken together, these data suggest that interventions are most likely to support stability when they operate within a broader, credible policy framework, particularly a clear commitment to the exchange rate regime and accompanying restrictions that shape expectations and reduce short-term FX demand. At the same time, the monthly intervention series and volatility in the official exchange rate cannot, on their own, establish a stabilising causal relationship, since interventions often respond to market stress and peg mechanics mechanically compress observed volatility.

4.2 Alternative explanations

Changes in exchange rate volatility during the war may also be driven by factors that shift independently of intervention volumes. These include adjustments to capital controls and currency restrictions that affect access to foreign exchange, changes in perceptions of reserve adequacy, announcements of external financing (for example, IMF, EU, or US support packages), and major battlefield or geopolitical developments that influence risk sentiment and expectations.

4.3 Limitations of the research and interpretation of results

These results should be read as descriptive patterns rather than causal effects. First, volatility is computed from the official exchange rate, so volatility under peg regimes is naturally very low because the rate is constrained by design. Second, interventions are reported at a monthly frequency while volatility is computed daily and then aggregated, meaning Figure 5 captures broad co-movement rather than precise timing or immediate effects.⁶ The post-month comparison in Table 2 is therefore interpreted as a structured descriptive check rather than evidence of timing or causality. Even with these limitations, the results still support a cautious

conclusion: stabilisation is more visible in the early stage of the war, while later regime changes and persistent shocks make it harder to sustain similarly clear outcomes. Because interventions often respond to market stress and are only observed monthly, the analysis cannot cleanly separate intervention effects from underlying shocks; doing so would require an event-study or regression design using higher-frequency information.

5. Acknowledgements

I express my sincere appreciation to Dr. Ceyhun Elgin and Dr. George Zifkos for their continuous support, mentorship, and detailed feedback throughout the development of this research. Their expertise and guidance helped shape the direction of the analysis and strengthened the final manuscript.

6. Conclusion

6.1 Summary of the study

The study examined whether foreign exchange interventions by the National Bank of Ukraine (NBU) were related to changes in volatility in the UAH/USD exchange rate in the pre-war period and during the war. Daily official exchange rate data from 2018 to 2026 provided context for baseline hryvnia movements, and these data were analysed alongside monthly net totals of NBU foreign exchange interventions. Volatility was measured using two approaches: the standard deviation of daily log returns within each exchange rate regime, and a 30-observation rolling standard deviation that was averaged at the monthly level to align with the frequency of the NBU's intervention reports.^{5,6} It was hypothesised that interventions would be associated with lower volatility in the short-run, but that this association would weaken as the war progressed.

6.2 Main findings

Measured volatility in the official exchange rate appears to be driven primarily by the exchange rate regime, with peg periods mechanically compressing day-to-day movements and therefore lowering the volatility statistic. Regime-based volatility is highest pre-war (0.0045), mechanically near zero under the two pegs (0.0005 and 0.0003), and higher under managed flexibility (0.0026). Within the two peg regimes, exchange rate volatility remained minimal and near zero. This resulted largely from the administrative constraint that the official exchange rate had to remain effectively constant and therefore could not respond to market pressures.^{7,8,11} The rise in exchange rate volatility in mid-2022, as measured by the rolling exchange rate standard deviation, could be explained by the adjustment from one fixed exchange rate level to another.⁸ In other words, the fixed exchange rate produced a sequence of near-zero daily return values. These values remained within the 30-observation window and, once the discrete exchange rate adjustment occurred, that single large change entered the window as an outlier and mechanically increased measured volatility. However, as the rolling window progressed, the measure returned

toward its normal level once the adjustment moved out of the 30-observation sample. Apart from the above, and following the shift to managed exchange rate flexibility in October 2023, exchange rate volatility became noticeable again, as expected when the official rate resumed more frequent day-to-day adjustment under continued central-bank management.^{9,12}

6.3 Interpretation relative to the hypothesis

When monthly exchange rate volatility was compared with monthly net intervention, the series showed broad co-movement: larger net FX sales often occurred in higher-volatility months. This is precisely the pattern expected under an endogenous policy reaction, since large FX sales are undertaken when exchange market pressure is already elevated, rather than providing clear evidence of stabilisation.⁶ Therefore, the hypothesis received only partial support. During the initial phase of the conflict, volatility in the official exchange rate was mechanically compressed under the peg, which limited how clearly any intervention–volatility pattern could be identified. Once volatility became more visible later in the sample, co-movement with net intervention remained uneven. This suggests that sustained shocks and evolving policy conditions reduced the extent to which intervention volumes alone could account for the volatility of outcomes observed in the official exchange rate.¹¹

6.4 Broader questions and limitations

These issues raised broader questions about how stabilisation should be assessed under crisis conditions. In particular, they highlighted the extent to which observed “stability” reflected the exchange rate regime itself rather than the marginal effects of intervention. In addition, because interventions were measured monthly and volatility was derived from daily returns, the analysis captured co-movements rather than precise timing, which limited how confidently the direction of influence could be inferred.⁶ Finally, under prolonged uncertainty, expectations and policy signalling may have played a larger role in shaping exchange rate behaviour than intervention quantities alone, which could further blur any simple intervention–volatility link.² For that reason, exchange rate volatility alone is an incomplete measure of intervention effectiveness. A fuller assessment would also use market indicators such as the cash/parallel-market premium and basic liquidity measures (e.g., bid–ask spreads or FX turnover), which this study does not systematically test.

6.5 Suggestions for future research

Future research could address these issues by separating regime effects from intervention effects more directly. Event-study methods around key policy changes, such as the July 2022 adjustment and the October 2023 transition to managed flexibility, could help identify short-run volatility responses more precisely.^{8,9} In addition, using higher-frequency intervention measures where available, and comparing official-rate volatility with market-rate volatility during peg periods, would offer a clearer benchmark for underlying market instability when the official series is mechanically constrained.¹¹ More generally, future work could estimate lagged

regressions or event studies around key policy episodes using market-rate volatility and higher-frequency intervention measures to better separate regime effects from intervention effects.

6.6 Policy implications: official vs market-rate stability

Stability in the official UAH/USD rate, especially during the peg periods, does not necessarily translate into stability in the exchange rates faced by households and firms. When access to FX is restricted, market conditions can shift toward cash or parallel markets, where premia and volatility may differ from the official series. This implies that evaluating intervention effectiveness using only official-rate volatility can overstate the degree of stability experienced in the broader economy, and that a fuller assessment should also consider market-rate indicators such as cash/parallel-rate movements and the size of the premium.

Works Cited

- (1) Dziamulych, M.; Krupiak, I.; Borysiuk, O.; Korobchuk, T.; Chyzh, N. Influence of the monetary policy of the NBU on the activity of banking institutions in Ukraine during the war. *Cadernos de Educação Tecnologia e Sociedade* 2024, 17 (SE3), 278–286. <https://doi.org/10.14571/brajets.v17.nse3.278-286>.
- (2) Holub, R.; Hlushchenko, O. The National Bank of Ukraine communication strategy optimization within the framework of impact on exchange rate expectations of economic agents. *Economies* 2017, 5 (3), 33. <https://doi.org/10.3390/economies5030033>.
- (3) Menkhoff, L. Foreign exchange intervention in emerging markets: a survey of empirical studies. *The World Economy* 2013, 36 (9), 1187–1208. <https://doi.org/10.1111/twec.12027>.
- (4) Tomilina, M. Monetary Policy in Ukraine during the War and Post-War Periods. Master's Thesis, National University of Kyiv-Mohyla Academy, Kyiv, Ukraine, 2023. <https://ekmair.ukma.edu.ua/items/62e88dda-8cc4-4601-bec2-aa9ae0417b00> (accessed January 5, 2026).
- (5) National Bank of Ukraine. Official hryvnia exchange rates (data). <https://bank.gov.ua/en/markets/exchangerate-chart> (accessed January 6, 2026).
- (6) National Bank of Ukraine. NBU currency interventions (data). <https://bank.gov.ua/en/markets/currency-interventions> (accessed January 6, 2026).
- (7) National Bank of Ukraine. How Ukraine's banking system and FX market will work from 24 February 2022 under martial law throughout Ukraine. News release, February 24, 2022. <https://bank.gov.ua/en/news/all/pro-robotu-bankivskoyi-sistemi-ta-valyutnogo-rinku-z-24-lyutogo-2022-roku-za-umovi-voyennogo-stanu-po-vsiy-teritoriyi-ukrayini> (accessed January 10, 2026).
- (8) National Bank of Ukraine. NBU fixes official UAH/USD exchange rate at a new level and takes additional measures to balance the FX market and support resilience of economy during the war. News release, July 21, 2022. <https://bank.gov.ua/en/news/all/nbu-zafiksuvav-ofitsiyinyi-kurs-grivni-do-dolara-ssha-nanovomu-rivni-ta-vjiv-nizku-dodatkovih-zahodiv-dlya-zbalansuvannya-valyutnogo-rinku-ta-pidtrimannya-stiykosti-ekonomiki-v-umovah-viyni> (accessed February 14, 2026).
- (9) National Bank of Ukraine. NBU introduces managed flexibility of exchange rate to strengthen resilience of the economy and the FX market. News release, October 2, 2023 (effective October 3, 2023). <https://bank.gov.ua/en/news/all/nbu-vprovadjuje-kerovanu-gnuchkist-obminnogo-kursu-scho-posilit-stiykist-valyutnogo-rinku-ta-ekonomiki> (accessed January 15, 2026).
- (10) Grui, A.; Aragon, N.; Faryna, O.; Krukovets, D.; Savolchuk, K.; Sulimenko, O.; Vdovychenko, A.; Zholud, O. Between Russian invasions: the monetary policy transmission mechanism in Ukraine in 2015–2021. NBU Working Paper 02/2023; National Bank of Ukraine: Kyiv, Ukraine, 2023.

https://bank.gov.ua/ua/file/download?file=WP_2023-02_Grui_et_al.pdf (accessed January 17, 2026).

- (11) International Monetary Fund. Ukraine: 2023 Article IV Consultation, Second Review under the Extended Arrangement under the Extended Fund Facility, and Requests for Modification of Performance Criteria and a Waiver of Nonobservance of Performance Criterion—Press Release; Staff Report; and Statement by the Executive Director for Ukraine. IMF Country Report No. 23/399; International Monetary Fund: Washington, DC, 2023.

<https://www.imf.org/en/publications/cr/issues/2023/12/11/ukraine-2023-article-iv-consultation-second-review-under-the-extended-arrangement-under-the-542297> (accessed January 19, 2026).

- (12) Reuters. Ukraine eases fixed exchange rate for first time during war; central bank introduces “managed flexibility”. October 2, 2023.

<https://www.reuters.com/markets/currencies/ukraines-central-bank-introduces-managed-flexibility-exchange-rate-2023-10-02/> (accessed January 25, 2026).

Attachment Insecurity in Bipolar Spectrum Disorders: A Systematic Review of Prevalence and Severity By Byron Kwon

Abstract

Bipolar spectrum disorders are characterized by profound mood instability that often leads to significant interpersonal dysfunction. Attachment theory provides a critical framework for understanding these relational challenges through the lens of internal working models. This systematic review examines the prevalence of various adult attachment styles in bipolar populations and evaluates the impact of attachment insecurity on relational outcomes. A comprehensive search identified 16 empirical studies (N = 21 to 161) utilizing validated assessment tools, including the Experiences in Close Relationships (ECR) scale and the Relationship Scales Questionnaire (RSQ). Findings indicate that attachment anxiety is highly prevalent among adults with Bipolar Disorder, with insecurity rates ranging from 49.7% to 89%. Conversely, levels of attachment avoidance appear heterogeneous across the literature. Furthermore, the data suggest a significant correlation between childhood trauma and adult depressive severity, mediated by attachment insecurity. The majority of individuals within the bipolar spectrum exhibit insecure attachment patterns which may impede healthy relational functioning. These results suggest that therapeutic interventions should include attachment-focused strategies to improve long-term psychosocial outcomes.

Introduction

Bipolar spectrum disorders are chronic, recurrent mood disorders characterized by episodes of mania, hypomania, and depression that substantially impair interpersonal functioning, quality of life, and long-term psychosocial outcomes. Beyond mood symptoms alone, individuals with bipolar disorder frequently experience persistent difficulties in close relationships, including heightened sensitivity to rejection, interpersonal instability, and reduced social support. These relational vulnerabilities have increasingly been examined through the lens of attachment theory, which provides a developmental framework for understanding how early relational experiences shape emotion regulation and interpersonal behavior across adulthood.

Attachment theory posits that individuals develop internal working models of self and others that guide expectations in close relationships. In adulthood, attachment is commonly conceptualized along two continuous dimensions: attachment anxiety, reflecting fear of abandonment and excessive reassurance-seeking, and attachment avoidance, reflecting discomfort with intimacy and emotional closeness. Elevated levels of either dimension indicate attachment insecurity, whereas low levels of both dimensions reflect secure attachment. Attachment insecurity has been robustly linked to affective dysregulation, stress reactivity, and vulnerability to mood disorders in the general population.

In bipolar disorder specifically, attachment insecurity has been implicated in illness courses, functional impairment, emotional regulation difficulties, and response to stress. Empirical studies suggest that adults with bipolar disorder may exhibit higher rates of insecure

attachment and elevated attachment anxiety and avoidance compared with healthy controls. Moreover, attachment insecurity has been associated with greater depressive symptom severity, reduced resilience, impaired social functioning, and the impact of childhood trauma in bipolar samples. However, findings across studies vary substantially in reported prevalence rates, severity estimates, and attachment profiles, reflecting differences in diagnostic subtypes, mood state, measurement tools, and analytic approaches.

Despite a growing literature examining attachment in bipolar disorder, no consensus has been reached regarding the prevalence of attachment insecurity or the typical severity of attachment anxiety and avoidance among adults with bipolar spectrum disorders. Individual studies often report heterogeneous outcomes, and attachment has been assessed using both categorical and dimensional models, further complicating synthesis. A systematic evaluation of the existing evidence is therefore needed to clarify overall patterns and provide clinically meaningful estimates.

Accordingly, the present study systematically reviews empirical research addressing the following question: Among adults diagnosed with bipolar spectrum disorders, what is the prevalence and typical severity of attachment insecurity, including attachment anxiety and attachment avoidance? By synthesizing findings across multiple studies, this review aims to characterize attachment profiles in bipolar disorder and inform future research and clinical interventions targeting interpersonal functioning in this population.

Method

Study design

This study employed a systematic review with quantitative data extraction to examine the prevalence and severity of attachment insecurity in adults diagnosed with bipolar spectrum disorders. The review focused on synthesizing both categorical attachment classifications (secure vs. insecure) and dimensional measures of attachment anxiety and avoidance across empirical studies.

Literature Search

A systematic literature search was conducted using PubMed, Google Scholar, and reference list screening to identify peer-reviewed empirical studies published in English. Search queries combined terms related to bipolar spectrum disorders (e.g., bipolar disorder, bipolar I, bipolar II, bipolar spectrum, cyclothymia) with terms related to attachment (e.g., attachment, attachment style, attachment insecurity, anxious attachment, avoidant attachment, adult attachment, ECR, AAI, ASQ, RSQ). Reference lists of relevant articles and review papers were manually examined to identify additional eligible studies. When multiple publications reported findings from the same dataset, the most comprehensive or statistically informative report was retained.

Eligibility Criteria

Studies were included if they met the following criteria: (a) the sample consisted of adults (≥ 18 years) diagnosed with a bipolar spectrum disorder, including bipolar I disorder, bipolar II disorder, cyclothymia, or bipolar spectrum disorder as defined by study authors; (b) attachment was assessed using a validated adult attachment measure, either categorically (e.g., secure vs. insecure) or dimensionally (attachment anxiety and/or avoidance); (c) the study employed an empirical cross-sectional or longitudinal design; and (d) sufficient data were reported to describe prevalence of attachment insecurity and/or mean levels of attachment anxiety and avoidance.

Studies were excluded if they focused exclusively on children or adolescents, did not report attachment outcomes separately for bipolar participants, relied solely on qualitative methods, or consisted of reviews, dissertations, or theoretical papers without extractable empirical data.

Study Selection

Titles and abstracts identified through the search were screened for relevance, followed by full-text review to confirm eligibility. Using this process, empirical studies examining attachment in adults with bipolar spectrum disorders were identified and included. Review articles and dissertations were retained for background context but excluded from quantitative synthesis.

Variables and Operational Definitions

Bipolar Spectrum Disorders. Bipolar spectrum disorders included bipolar I disorder, bipolar II disorder, cyclothymic disorder, and bipolar spectrum or bipolar-related diagnoses as defined by DSM or ICD criteria within each study.

Attachment Insecurity. Attachment insecurity was defined either categorically (secure vs. insecure attachment classifications) or dimensionally, based on elevated levels of attachment anxiety and/or attachment avoidance. Attachment was assessed using validated self-report or interview-based instruments, including the Experiences in Close Relationships scales (ECR, ECR-II, ECR-R), Attachment Style Questionnaire (ASQ), Relationship Scales Questionnaire (RSQ), Adult Attachment Scale (AAS), and related measures.

Data Extraction

From each included study, data were systematically extracted on publication year, country, diagnostic composition of the bipolar sample, sample size, age characteristics, attachment measure used, availability of attachment classification, prevalence of insecure attachment (when reported), and mean levels and standard deviations of attachment anxiety and avoidance. When multiple attachment dimensions or subgroups were reported, all relevant bipolar-specific data were extracted.

Data Synthesis

Given substantial heterogeneity in diagnostic composition, attachment measures, and analytic strategies, findings were synthesized using a narrative approach supplemented by descriptive quantitative reporting. Prevalence rates of attachment insecurity and mean severity estimates for attachment anxiety and avoidance were summarized across studies. No pooled meta-analytic estimates were calculated due to variability in measurement scales and study designs.

Methodological Considerations

Most included studies relied on cross-sectional designs and self-report attachment measures, limiting causal inference and raising the possibility of mood-state or recall bias. Additional limitations include small sample sizes in several studies, variability in diagnostic criteria, and inconsistent reporting of attachment classifications. These issues are considered in the Discussion.

Table 1. Study Characteristic and Data Extraction for Bipolar Disorder and Attachment Insecurity

Author (Year)	Sample (N)	Measure	Key Findings^b	Effect Size (d)^a
Gilbert et al. (2007)	40 (B)	AAS	AI is associated with depression severity in BD.	N/R
Morriss et al. (2009)	107 (B), 41 (C)	RSQ	Higher AI in BD-I vs HC; fluctuates with mood.	-0.63 (Dep)
Kökçü et al. (2010)	44 (B), 84 (C)	AAS	BD showed higher AA and AV than HC.	N/R
Harnic et al. (2014)	90 (B), 86 (C)	ECR	BD and cyclothymia showed higher AI than HC.	N/R
Picardi et al. (2019)	21 (B), 42 (C)	ECR	BD showed higher AA/AV than all controls.	N/R
Wagner-Skacel (2020)	46 (B)	ECR-RS	AI is associated with higher stress/symptom severity.	N/R
Karaş et al. (2021)	63 (B), 63 (C)	ECR-II	BD-I showed lower AV but similar AA vs HC.	N/R
Çıtak et al. (2020)	110 (B)	ECR-R	CT associated with AA/AV and lower resilience.	N/R
Wrobel et al. (2022)	143 (B)	ECR	AI mediated the link between CT and depression.	N/R
Craba et al. (2023)	55 (B), 60 (C)	ECR	Fearful attachment (60%) is the most common style in BD.	N/R

Kartal et al. (2023)	63 (B), 61 (C)	RSQ	AI and temperament linked to functional impairment.	N/R
Rahmatinejad (2018)	41 (B)	Hazan & Shaver AAS	BD-II and BPD shared high AI; BPD had higher AV.	N/R
Morán-Kneer (2022)	76 (B)	ECR	AA moderated the link between CT and social cognition.	N/R
Fuhr et al. (2017)	61 (B), 60 (C)	RSQ	AI linked to DEP via self-esteem/attitudes.	N/R
Kurt et al. (2025)	100 (B), 100 (C)	ECRI-II	AI linked to CT, dissociation, and alexithymia.	N/R

Note. N = total sample size; B = Bipolar spectrum group; C = Control group (healthy or clinical); AA = Attachment Anxiety; AV = Attachment Avoidance; AI = Attachment Insecurity; CT = Childhood Trauma; DEP = Depressive symptoms; NR = Not Reported.

^a Effect sizes are reported as Cohen's d where available; negative values indicate lower scores in the bipolar group relative to controls or mood states.

^b Prevalence percentages refer to the proportion of the bipolar sample classified as "Insecure" (Anxious, Avoidant, or Fearful) versus "Secure."

Results

Study Characteristics

A total of 18 empirical studies and reviews were identified. Sample sizes for bipolar groups ranged from N=21 to N=161. The majority of studies utilized the Experiences in Close Relationships (ECR) scale or the Relationship Scales Questionnaire (RSQ). Most samples consisted of euthymic (remitted) outpatients, though some studies included participants in active manic or depressive states.

Prevalence of Attachment Insecurity

Across studies providing categorical data, adults with bipolar spectrum disorders exhibited high rates of attachment insecurity. Prevalence rates for insecurity ranged from 49.7% (Greenberg, 2014) to as high as 89% (Harnic et al., 2014). Comparison to Controls: In studies with healthy control groups, the bipolar group consistently showed significantly higher insecurity. For instance, Morriss et al. (2009) found 78% insecurity in Bipolar I participants compared to 32% in healthy controls.

Categorical Profiles. Where specific styles were reported, fearful-avoidant and anxious-preoccupied styles were most prevalent (Craba et al., 2023).

Severity of Attachment Anxiety and Avoidance. Dimensional scores further clarified the nature of this insecurity.

Attachment Anxiety. Attachment Anxiety was consistently elevated across the spectrum. Studies like Picardi et al. (2019) reported high mean anxiety scores ($M=83.6$), and Kurt et al. (2025) found significantly higher anxiety in BD participants ($M=67.19$) compared to controls.

Attachment Avoidance. The attachment avoidance style displayed more heterogeneous results. While Harnic et al. (2014) and Picardi et al. (2019) found elevated avoidance, Karaş et al. (2021) actually reported lower attachment avoidance in BD-I patients compared to controls. This suggests that while anxiety (fear of abandonment) is a core trait, avoidance (discomfort with intimacy) may vary based on the specific bipolar subtype or cultural context.

Clinical and Functional Correlates

Attachment insecurity was significantly associated with worse clinical outcomes. High attachment anxiety moderated the relationship between childhood trauma and depression severity (Wrobel et al., 2022; Çıtak et al., 2020). Additionally, secure attachment was a significant predictor of better social functioning and resilience (Kartal et al., 2023).

Conclusion

This systematic review confirms that attachment insecurity is a hallmark of the bipolar spectrum, with attachment anxiety serving as a particularly consistent clinical feature. The high prevalence of fearful and preoccupied attachment styles suggests that individuals with bipolar disorder navigate a "double burden" of desiring closeness while simultaneously fearing rejection or intimacy. Crucially, the evidence highlights a trauma-attachment-depression pathway, where insecure attachment mediates the long-term impact of childhood trauma on adult symptom severity.

These findings have significant implications for clinical practice. Current treatment models for bipolar disorder often prioritize pharmacological stabilization and symptom management. However, the prevalence of attachment insecurity suggests that clinicians should integrate attachment-focused interventions—such as Interpersonal and Social Rhythm Therapy (IPSRT) or family-focused therapy, to address the underlying relational vulnerabilities of the disorder. By fostering more secure internal working models, therapeutic outcomes and long-term psychosocial functioning may be significantly improved.

Limitations

Several methodological limitations must be considered when interpreting these findings. First, the majority of included studies utilized cross-sectional designs, which preclude the ability to draw causal inferences regarding the development of attachment insecurity over the course of the illness. Second, the reliance on self-report measures, such as the ECR and RSQ, introduces

the risk of mood-state bias; a participant's current depressive or manic symptoms may skew their perception of relational history and current attachment security.

Furthermore, there was significant heterogeneity in the clinical characteristics of the samples. Variability in diagnostic subtypes (Bipolar I vs. Bipolar II), current mood states (euthymic vs. symptomatic), and cultural backgrounds makes it difficult to generalize a single attachment profile to the entire bipolar spectrum. Finally, many studies were limited by small sample sizes, which may reduce the statistical power of the reported correlations. Future research should prioritize longitudinal designs and larger, multi-site cohorts to better understand how attachment dimensions evolve across different mood episodes and whether targeted attachment interventions can alter the clinical trajectory of the disorder.

Acknowledgements

I would like to thank Jaimie Park, Min Kyung Kang, and Prestige Institute for their guidance and support in my research process.

Works Cited

- Ambra Craba, Marano, G., Kotzalidis, G. D., Avallone, C., Francesco Maria Lisci, Maria Luigia Crosta, Callea, A., Monti, L., Domenico De Berardis, Lai, C., Balocchi, M., Sessa, I., D. Harnic, Sani, G., & Mazza, M. (2023). Resilience and Attachment in Patients with Major Depressive Disorder and Bipolar Disorder. *Journal of Personalized Medicine*, 13(6), 969–969. <https://doi.org/10.3390/jpm13060969>
- Celaleddin Kartal, & Esat Fahri Aydın. (2023). Functional Impairment, Attachment, and Affective Temperaments in Euthymic Patients with Bipolar Disorder. *Psychiatry and Clinical Psychopharmacology*, 33(4), 299–308. <https://doi.org/10.5152/pcp.2023.22597>
- Citak, C., & Erten, E. (2021). Impact of Childhood Trauma and Attachment on Resilience in Remitted Patients with Bipolar Disorder. *Journal of Affective Disorders*, 280, 219–227. <https://doi.org/10.1016/j.jad.2020.11.025>
- Clarkson, K., & Hosang, G. (2012). Childhood attachment and parental bonding in bipolar disorder. *International Clinical Psychopharmacology*, 28, e45. <https://doi.org/10.1097/01.yic.0000423320.63304.74>
- Fuhr, K., Reitenbach, I., Kraemer, J., Hautzinger, M., & Meyer, T. D. (2017). Attachment, dysfunctional attitudes, self-esteem, and association to depressive symptoms in patients with mood disorders. *Journal of Affective Disorders*, 212, 110–116. <https://doi.org/10.1016/j.jad.2017.01.021>
- Gilbert, P., McEwan, K., Hay, J., Irons, C., & Cheung, M. (2007). Social rank and attachment in people with bipolar disorder. *Clinical Psychology & Psychotherapy*, 14(1), 48–53. <https://doi.org/10.1002/cpp.508>
- Greenberg, S., Rosenblum, K. L., McInnis, M. G., & Muzik, M. (2014). The role of social relationships in bipolar disorder: A review. *Psychiatry Research*, 219(2), 248–254. <https://doi.org/10.1016/j.psychres.2014.05.047>
- Greenberg, S., Stein, C., Carels, R., & Zickar, M. (2014). ATTACHMENT STYLE, PERCEIVED LIFE EVENTS, AND PSYCHOLOGICAL WELL- BEING IN ADULTS COPING WITH BIPOLAR DISORDER: A LONGITUDINAL STUDY Committee.
- Harnic, D., Pompili, M., Innamorati, M., Erbutto, D., Lamis, D. A., Bria, P., Girardi, P., & Janiri, L. (2014). Affective temperament and attachment in adulthood in patients with Bipolar Disorder and Cyclothymia. *Comprehensive Psychiatry*, 55(4), 999–1006. <https://doi.org/10.1016/j.comppsy.2013.12.006>
- Karaş, H., Çarpar, E., Küçükgöncü, S., & Kaşer, M. (2021). Associations of Attachment and Coping Styles With Social Functioning in Patients With Bipolar Disorder I. *Journal of Nervous & Mental Disease*, Published Ahead of Print. <https://doi.org/10.1097/nmd.0000000000001352>
- Kökçü, F., & Kesebir, S. (2010). [The relationship between attachment style, and temperament, personality and bipolar symptoms: a controlled study on bipolar patients and their children]. *Türk Psikiyatri Dergisi = Turkish Journal of Psychiatry*, 21(4), 309–318. <https://pubmed.ncbi.nlm.nih.gov/21125506/>

- Kouros, I., Isaksson, M., Ekselius, L., & Ramklint, M. (2024). A cluster analysis of attachment styles in patients with borderline personality disorder, bipolar disorder and ADHD. *Borderline Personality Disorder and Emotion Dysregulation*, 11(1).
<https://doi.org/10.1186/s40479-024-00271-2>
- Kurt, M., & Gülpek, D. (2025). THE RELATIONSHIP BETWEEN CHILDHOOD TRAUMA, DISSOCIATION, ATTACHMENT AND ALEXITHYMIA IN BIPOLAR AFFECTIVE DISORDER PATIENTS. *Archives of Neuropsychiatry*.
<https://doi.org/10.29399/npa.28924>
- Morán-Kneer, J., Ríos, U., Costa-Cordella, S., Barría, C., Carvajal, V., Valenzuela, K., & Wasserman, D. (2022). Childhood trauma and social cognition in participants with bipolar disorder: The moderating role of attachment. *Journal of Affective Disorders Reports*, 9, 100359. <https://doi.org/10.1016/j.jadr.2022.100359>
- Morriss, R. K., Gucht, E., Lancaster, G., & Bentall, R. P. (2009). Adult attachment in bipolar 1 disorder. *Psychology and Psychotherapy: Theory, Research and Practice*, 82(3), 267–277.
<https://doi.org/10.1348/147608309x415309>
- Picardi, A., Pallagrosi, M., Fonzi, L., Martinotti, G., Caroppo, E., Meldolesi, G. N., Di Gennaro, G., De Risi, M., & Biondi, M. (2019). Attachment in Patients with Bipolar and Unipolar Depression: A Comparison with Clinical and Non-clinical Controls. *Clinical Practice and Epidemiology in Mental Health : CP & EMH*, 15, 143–152.
<https://doi.org/10.2174/1745017901915010143>
- Rahmatinejad, P., Khosravi, Z., & Mohammadi, S. D. (2018). Comparing Attachment Style, Quality of Intimate Relationship and Anger Experience in Patients with Borderline Personality and Bipolar-II Disorders. *Iranian Journal of Psychiatry and Behavioral Sciences*, 12(2). <https://doi.org/10.5812/ijpbs.10720>
- Wagner-Skacel, J., Bengesser, S., Dalkner, N., Mörtl, S., Painold, A., Hamm, C., Pilz, R., Rieger, A., Kapfhammer, H.-P., Hiebler-Ragger, M., Jauk, E., Butler, M. I., & Reininghaus, E. Z. (2020). Personality Structure and Attachment in Bipolar Disorder. *Frontiers in Psychiatry*, 11. <https://doi.org/10.3389/fpsy.2020.00410>
- Wrobel, A. L., Russell, S. E., Jayasinghe, A., Lotfaliany, M., Turner, A., Dean, O. M., Cotton, S. M., Diaz-Byrd, C., Yocum, A. K., Duval, E. R., Ehrlich, T. J., Marshall, D. F., Berk, M., & McInnis, M. G. (2022). Attachment insecurity partially mediates the relationship between childhood trauma and depression severity in bipolar disorder. *Acta Psychiatrica Scandinavica*, 145(6). <https://doi.org/10.1111/acps.13419>

Urinal as Art: Investigating Context through Marcel Duchamp's 1917 "Readymade", *Fountain* By San Yun

Abstract

This paper examines the role of context in shaping the meaning of Marcel Duchamp's "readymades", with particular focus on *Fountain* (1917). By situating Duchamp's work within the social, industrial, and cultural conditions of the early twentieth century, the study demonstrates how World War I, the Second Industrial Revolution, and the rise of avant-garde movements such as Dada informed his rejection of traditional artistic practices. Through analysis of *Nude Descending a Staircase, No. 2* (1912), the paper traces Duchamp's growing dissatisfaction with painting and his critique of "retinal art," which ultimately led to the adoption of the "readymade". The paper argues that Duchamp's selection of mass-produced objects and their presentation within institutional settings redefined art as an intersection of concept, choice, and context rather than craftsmanship or aesthetic form. In doing so, *Fountain* emerges not as a mere provocation but as an intellectual response to the shifting cultural values of modernity, establishing a framework that continues to shape contemporary artistic practice.

Introduction

Marcel Duchamp (1887-1968) is regarded as one of the most consequential and controversial twentieth-century artists, expanding the definition of art to something far broader. As an artist most associated with Dada, Duchamp did not concern himself with the traditional rendering of talent, beauty, and craftsmanship but instead with concept, choice, and situation. This paper aims to demonstrate that the importance of context in Duchamp's "readymades" is not incidental, but central to a redefinition of art that prioritizes concept and placement over craftsmanship.

Historical and Cultural Context of the Early Twentieth Century

The 1910s were a period of unprecedented crisis and change. As the First World War broke out in 1914, decades, even centuries, of political affiliations and cultural sensibilities came crashing down. With many casualties, cities ravaged by bombs, the most powerful institutions—government, religion, and art—became questionable.¹⁰ With such international strife, many artists moved to New York to give their world a safer haven: an intellectual, cultured environment that at least delayed the necessary recess in avant-garde development. As Duchamp moved to New York in 1915, he noted that he was "able to work again."¹¹

Simultaneously, the Second Industrial Revolution changed the quality of life. Where there was once creation and limited access to things, now, there were manufactured goods at every turn, and the practicality of such and the collection that came about from mass production, which exceeded the value of what was handmade. The idea of an object's inherent individuality

¹⁰ Encyclopaedia Britannica, "World War I," accessed December 2025, <https://www.britannica.com/event/World-War-I>.

¹¹ Calvin Tomkins, *Duchamp: A Biography* (New York: Henry Holt and Company, 1996), 186.

had eroded, even though uniqueness had traditionally been regarded as a defining principle of fine art. Duchamp was more open-minded to these changes and developments because they facilitated the tangible and intangible creations of his later works.¹²

The era was also marked by decisive breaks from academic conventions, as artists increasingly rejected traditional standards of representation and meaning. Movements such as Cubism fragmented form and perspective, Futurism emphasized speed and mechanization, and Expressionism prioritized emotional intensity over realism. Within this broader shift, Dada emerged as the most radical response, rejecting not only aesthetic norms but the very idea that art required coherence, skill, or purpose. Rather than creating beauty, Dada artists embraced irrationality, irony, and disruption, reflecting a world that no longer seemed governed by reason.¹³ It is within this context that Duchamp's work becomes legible as a logical extension of a broader collapse of artistic and cultural conventions.

Nude Descending a Staircase and the Rejection of “Retinal Art”

Nude Descending a Staircase, No. 2 (1912) (Fig. 1), was Duchamp's second exploration into movement in paint, where his earlier version was completed in 1911. While a “realistic” nude would celebrate the feminine form, instead, Duchamp took the opportunity to attempt to render an idea of movement.¹⁴ Rather, the abstraction comes in a confusing array of overlapping shapes to suggest a body descending the stairs. Duchamp became frustrated with the realism of what he could paint and challenged himself to surpass traditional subject matter and painting techniques, emphasizing how “retinal art” gave him zero interest.¹⁵

In *Nude Descending a Staircase, No. 2*, Duchamp portrays a dissolution of the human figure into rhythmic, geometric forms that convey motion rather than fixed identity. The piece is a large, vertical canvas, composed of a patterned, geometric mass that fractures the human form into sections. Duchamp breaks the body into several overlapping forms by using the logic of chronophotography: a branch of instantaneous photography that captures phases of movement.¹⁶ The sections are diagonal and fluid, running from top left to bottom right, as if it's falling. The colors are not distracting with an overwhelming vibrancy, as the browns and ochres are relatively monochromatic, paying tribute to the rhythm and movement of the depicted figure. Thus, there are cubist elements in the sectioned-off pieces; there are futurist elements in the velocity and motion, and it has a chronophotographic quality to it as the viewer sees the figure in multiple moments simultaneously.

When *Nude Descending a Staircase, No. 2* was submitted in 1912 at the Salon des Indépendants in Paris, it baffled many and received amazing criticism from peer review, as

¹² Calvin Tomkins, *Duchamp: A Biography* (New York: Henry Holt and Company, 1996), 186.

¹³ Museum of Modern Art, “Dada,” collection essay, accessed March 2025, https://www.moma.org/learn/moma_learning/themes/dada/.

¹⁴ Kenneth Clark, *The Nude: A Study in Ideal Form* (Princeton: Princeton University Press, 1956), 1–5.

¹⁵ Calvin Tomkins, *Duchamp: A Biography* (New York: Henry Holt and Company, 1996), 118–120.

¹⁶ *Scientific American*, “Chronophotography,” *Scientific American* 75, no. 19 (November 7, 1896): 344, <https://www.jstor.org/stable/10.2307/26118672>.

well.¹⁷ Conservatives were not fans of the unrealistic abstraction of the nude, something already tied to the figuratively ideal female nude and, even, implicitly, spiritual features. But even Cubist painters, who one would think would appreciate disassociation, discarded it, going too far, they believed, because it was not abstract enough.¹⁸ Thus, the softened edges and industrialized essence of the figure inside *Nude Descending a Staircase* became an assault to many and a confused response to even more, indicating how far some art critics were from such an avant-garde, radical piece in this era.¹⁹



Fig 1: Marcel Duchamp, *Nude Descending a Staircase, No. 2*, 1912, Oil on canvas, 147 x 89.2 cm.
Philadelphia Museum of Art.

Image source: Cleveland Museum of Art

Created before the First World War, *Nude Descending a Staircase* is ahead of its time, a product of the technological revolution with a new approach to time, motion, and the human form. Therefore, the justification for movement rather than form comes from an allure with machines and modernity, which challenged traditional modes of representation. In addition, the fact that the work was so poorly received played into his later departure from painting and critical components that led him to explore more “readymade” avenues. Thus, this work reveals the history and context that pushed Duchamp to reject painting as a meaningful artistic practice.

20

¹⁷ Cleveland Museum of Art, “Infamy and Influence,” accessed December 2025, <https://www.clevelandart.org/articles/infamy-and-influence>.

¹⁸ Calvin Tomkins, *Duchamp: A Biography* (New York: Henry Holt and Company, 1996), 109–112.

¹⁹ Tomkins, *Duchamp: A Biography*, 110–115.

²⁰ Marcel Duchamp, *The Essential Writings of Marcel Duchamp*, ed. Michel Sanouillet and Elmer Peterson (London: Thames & Hudson, 1975), 138.

Dada and the Emergence of the “Readymade”

Duchamp’s departure from the canvas to his “readymades” came in Dada, an anti-art movement composed of a series of disconnected, often nonsensical pieces that valued absurdity over reason and made aesthetic gestures and ironic anti-establishment statements. To the Dadaists, a civilization that could support so much killing on the battlefields of World War I did not deserve reason or aesthetics, but instead, since the reality was a lamentable one, culturally created, used irony, comedy, and challenge to expose it.²¹

Duchamp’s emergence into the “readymade,” or mass-produced everyday objects selected by the artist for a new function as art, comes from this mentality. The idea of selecting something and calling it art, rather than creating something aesthetically appealing, was revolutionary. This act of choice, rather than physical creation, became an artistic gesture. Duchamp criticized the belief that art’s worth resided in craftsmanship or aesthetic appeal by eliminating the artist’s manual talent from the process.²²



Fig 2: Marcel Duchamp, *Fountain*, 1917 (replica 1964), Porcelain urinal, 36 × 48 × 61 cm. Philadelphia Museum of Art.

Image source: Alfred Stieglitz, photograph of *Fountain*, 1917.

Fountain (1917): Context and Controversy

Fountain (1917) (Fig. 2) is a mass-produced white porcelain urinal purchased by Duchamp from J. L. Mott Iron Works in New York City. When displaying the piece, Duchamp

²¹ Museum of Modern Art, “Dada.”

²² de Duve, *The Readymade and the Tube of Paint*, 89–94.

placed it on its side, signed it with the name “R. Mutt 1917,” and anonymously entered it into the Society of Independent Artists exhibition. Duchamp did not create it or sculpt it; he merely took it out of its functional space and placed it in a gallery without any elaborate trickery. According to William A. Camfield, the “readymade” was neither a practical joke nor a nihilist comment, but a constructed appeal against antiquated ideas of art and artistic creation.²³

Visually, *Fountain* (1917) presents itself as a stark, industrial object, whose smooth porcelain surface and utilitarian form contrast sharply with traditional expectations of sculpture. Removed from its functional setting and placed on its side, the urinal’s curved symmetry becomes newly visible, encouraging the viewer to perceive it as a sculptural form rather than a practical object. This reorientation does not alter the work materially, but instead reframes its meaning, emphasizing presentation over construction. The addition of the signature “R. Mutt 1917” further complicates this transformation, introducing authorship into an object that Duchamp did not physically create. As a result, the work resists interpretation based on craftsmanship or aesthetic refinement, directing attention instead to the conditions under which it is presented and recognized as art.

The reaction to *Fountain* was immediate and highly controversial, as the work was rejected despite the Society’s stated commitment to exhibiting all submissions.²⁴ Camfield notes that although the Society of Independent Artists promoted an open, non-juried exhibition, its board ultimately disregarded this principle in order to exclude a work they deemed inappropriate for display. Members of the board criticized *Fountain* as lacking artistic value, purpose, and legitimacy, revealing the limits of institutional openness when confronted with a challenge to established definitions of art.

Camfield examines *Fountain* in relation to the political, industrial, and cultural conditions of 1917, emphasizing how the work reflects the material realities of modern life. By introducing a mass-produced object associated with sanitation, industry, and the human body into the gallery space, Duchamp disrupts bourgeois expectations of propriety and artistic value. Rather than deriving meaning from craftsmanship or aesthetic form, *Fountain* demonstrates that artistic status is contingent upon context, presentation, and institutional framing. In this way, the work asserts itself as art not through what it is, but through where and how it is displayed, reinforcing Duchamp’s gesture as an intellectual response to the newfound modernity rather than a purely reactionary provocation.

Over a century after it was denied, *Fountain* stands as one of the most pivotal pieces of modern art because it exists with meaning inseparably bound to its surroundings. Duchamp took an item created in the same fashion as the rest of mass production, threw it in the trash, and then placed it in an exhibition, creating a system through which art could be viewed as valuable not for its appearance or construction but for placement, intended meaning, and institutional validation. Such logic is aligned with much of the art produced today—especially performance, multimedia, digital, and even the use of found objects in the artwork. Duchamp’s concept of

²³ William A. Camfield, “Marcel Duchamp’s *Fountain*: Its History and Aesthetics in the Context of 1917,” *Art Journal* 44, no. 1 (Spring 1984): 64–66.

²⁴ Camfield, “Marcel Duchamp’s *Fountain*,” 71–73.

questioning the nature of art and what art needs to be in the modern day holds some meaning in a society that is rapidly moving away from manufactured items and into the world of computers, digital art, and value. As a result, early twentieth-century creators should not be distinguished from their predecessors or successors, but instead, *Fountain* foreshadows an appropriate sensibility in our contemporary world, where social systems supply value instead of sites of production and aesthetics. By maintaining context as the most critical component of art value, Duchamp forever changed the way art is created, experienced, and valued by providing a means through which constraints uphold the piece forevermore.

Works Cited

- Britannica, T. Editors of Encyclopaedia. (2025). World War I. Encyclopaedia Britannica.
<https://www.britannica.com/event/World-War-I>
- Camfield, W. A. (1984). Marcel Duchamp's Fountain: Its history and aesthetics in the context of 1917. *Art Journal*, 44(1), 64–71.
- Clark, K. (1956). *The nude: A study in ideal form*. Princeton University Press.
- Cleveland Museum of Art. (2025). Infamy and influence.
<https://www.clevelandart.org/articles/infamy-and-influence>
- de Duve, T. (1996). *The readymade and the tube of paint*. MIT Press.
- Duchamp, M. (1975). *The essential writings of Marcel Duchamp* (M. Sanouillet & E. Peterson, Eds.). Thames & Hudson.
- Museum of Modern Art. (2025). Dada.
https://www.moma.org/learn/moma_learning/themes/dada/
- Philadelphia Museum of Art. (2025). Marcel Duchamp: Nude descending a staircase, no. 2.
<https://www.philamuseum.org/collection/object/51449>
- Scientific American. (1896). Chronophotography. *Scientific American*, 75(19), 344.
<https://www.jstor.org/stable/10.2307/26118672>
- Tomkins, C. (1996). *Duchamp: A biography*. Henry Holt and Company.

Scoliosis By Jingyi Huang

Abstract

This review is about scoliosis, a medical condition caused by lateral curvatures of the spine. This article introduces basic information about scoliosis, including its prevalence and its physical, mental, and psychological impacts on patients, especially adolescents. Specifically, this review discusses the physiological effects of scoliosis, including clinical diagnosis, the symptoms, treatments including both surgical and non-surgical interventions, and the prevention strategies of scoliosis implemented by different countries. The findings suggest that scoliosis management strategies should always be modified based on the characteristics of patients.

Introduction

Scoliosis is a medical condition characterized by a sideways curvature of the spine. Scoliosis usually develops during the growth spurt around puberty and can be manifested as a 'C' or 'S' shaped curve (Begum, 2024). Symptoms of scoliosis vary among patients, and potential symptoms include uneven shoulders, hips, back pain, and breathing difficulties in severe cases (Begum, 2024).

Currently, the prevalence of scoliosis is 0.47% to 5.2% of the population worldwide (Wikipedia, 2025). According to the National Scoliosis Foundation, about seven million people are diagnosed with scoliosis in the United States alone (Nalda, 2024). Among all types of scoliosis, idiopathic scoliosis is the most prevalent, but its cause remains unknown and may have hereditary links (Begum, 2024). Idiopathic scoliosis is common among adolescents: it occurs around the age of 11 to 18. It has a self-reported prevalence of 2-3% globally and a higher prevalence of 8.3-9.3% reported after X-ray screening (Zhang et al., 2024).

Due to the high incidence of scoliosis, it is crucial to understand the impacts on patients. This article discussed the outcome of scoliosis, intervention, and policies of scoliosis management

Methodology

This project is a literature review. It began with searching related literature with keywords, including scoliosis, idiopathic scoliosis, and prevention strategies. After locating articles of interest, the author read through each and summarized the key information, such as the specific cause of developmental scoliosis, its corresponding prevention, and the impacts on adolescents. A total of 15 articles are included in this literature review. Among these 15 articles, 12 are peer-reviewed literature published from 2014 to 2024, and three articles are from websites that contain educational information about scoliosis. The article screening began in mid-2024, and the drafting began in April 2025.

The findings include pathology, outcomes, intervention options, and prevention of scoliosis. These components are chosen because they can provide the general public with essential information regarding scoliosis. When drafting the article, medical jargon or

terminology is purposely removed since the audience who are not experts within the field may not understand the terminology.

Result

Pathology

There are two main types of scoliosis: idiopathic scoliosis and developmental scoliosis. Idiopathic scoliosis is genetically predisposed. Calmodulin is a calcium-binding receptor protein that regulates contractile properties of muscles and its increase in quantity is related to asymmetrical distribution of paraspinal muscles, which eventually leads to idiopathic scoliosis (Lataski et al., 2017). In contrast, developmental scoliosis is caused by postural dysfunction, which has been shown to be related to the lack of physical activities (Salsali et al., 2023).

Outcome of scoliosis

The outcome of scoliosis includes both physical and mental illness. The most eminent effect is the visible skeletal deformities, which include an uneven slanting of the shoulders and hip and pelvis alignment caused by uneven load of the spine that potentially leads to unequal lengths of legs. (NY Spine & Orthopedics (n.d.)). The same literature found that, in chronic cases, scoliosis prevents the rib cage from expanding fully, which limits cardiac output and respiratory capacity. Patients may experience symptoms of difficulty in deep breaths. Neurological deficits may also be found since the uneven compression of the spine can interrupt signals from the body to the brain, which contributes to a potential radiculopathy or neurological incontinence (NY Spine & Orthopedics (n.d.)).

The impacts of scoliosis vary between male and females as well. (NY Spine & Orthopedics (n.d.)) discussed that scoliosis compromises the female reproductive system as adolescent females with scoliosis are more likely to experience their first menstrual cycle later than their peers or irregular cycle. Some researchers now believe that low levels of female hormones—particularly progesterone—may cause scoliosis, and this may explain why adolescent idiopathic scoliosis is more common in girls than in boys (NY Spine & Orthopedics (n.d.)). Additionally, Glavas et al. (2023) found that eminent psychological consequences is more common in girls with scoliosis. The author mentioned that the prevalence of scoliosis escalated markedly in girls from 5.8% to 8.3%, while a modest decrease was noted in boys, from 3.8% to 3.2%. Girls with scoliosis are more likely to encounter school violence due to their ‘strange body posture’, and this can contribute to mental stress and loss of self-esteem (Glavas et al., 2023). Another contributing factor to mental stress among girls is that they are less willing to straighten their backbone due to their growing breasts (NY Spine & Orthopedics (n.d.)). Reduced self-esteem would lead to increased anxiety and depression from negative body image.

Diagnosis and management

The standard criteria of diagnosing scoliosis is Cobb angles greater than 10 degrees(Horne, 2014). An angle of trunk rotation greater than 5 degrees in the thoracic or lumbar segment of the spine is used as the criterion to identify suspected scoliosis(Zhang et al., 2024). The diagnosis has to be made by a physician. While symptoms of scoliosis vary among patients, there are only two main types of treatment: conservative approaches or surgical options. The comprehensive scoliosis care model includes education, physical therapy, observation, psychological support, bracing, and surgery. (Fan et al., 2020)

Conservative Approach

For patients who present with a low Cobb angle, conservative methods are the first line of treatment. The most commonly used conservative approach is called Physiotherapy Scoliosis Specific Exercises (PSSE), which aims to achieve prevention of surgery, restore posture and deformity, and improve respiratory capacity, strength, balance, and neuromuscular coordination so as to improve quality of life (QoL) (Fan et al., 2020). There are different approaches to PSSE. Fan et al., (2020) and Berdichevsky et al(2016) discussed approaches of PSSE in different countries: Lyon Approach in France combines PSSE with bracing and casting, focusing on 3D mobilization of the spine and patient education; Schroth Method in Germany emphasizes 3D postural correction, corrective breathing, and correction of postural perception; Scientific Exercise Approach to Scoliosis (SEAS) in Italy focuses on active self-correction and task-oriented exercises to improve spinal stability; Barcelona Scoliosis Physical Therapy School (BSPTS) in Spain is based on the principles of Katharina Schroth, focusing on muscle activation and breathing techniques; Dobomed Approach in Poland addresses both trunk deformity and respiratory function impairment, focusing on 3D auto-correction and asymmetrical breathing; Functional Individual Therapy of Scoliosis (FITS), another school in Poland, implemented a comprehensive method that includes various physiotherapeutic techniques for scoliosis correction; Side Shift Approach in United Kingdom is based on the theory that lateral movements can stabilize flexible curves and correct lateral deviation of the trunk. Although different methods are utilized in practice, PSSE is found to have a significant effect on reducing the Cobb angle in non-progressive scoliosis, which ultimately improves back asymmetry, muscle imbalance, and related pain(Fan et al., 2020; Berdichevsky et al., 2016).

Another conservative method, called Scoliosis-specific exercise (SSE), is prescribed during pubertal growth to slow progression of scoliosis(Fan et al., 2020). However, its effects are unclear, since there is insufficient evidence to prove that SSE, with or without other concurrent treatments, can reduce Cobb angle, improve trunk balance, or QoL (Fan et al., 2020). Thus, more evidence is needed in the future.

Surgical Approach

Different from conservative methods, surgery is suitable for more serious scoliosis with larger Cobb angles, and patients with a Cobb angle larger than 45-50 degrees are recommended for surgery (Koucheki et al.,2021).

Koucheki et al.(2021) conducted a meta-analysis comparing Enhanced Recovery after Spine Surgery (ERAS) with traditional discharge (TD) pathways for adolescent idiopathic scoliosis. The researchers included 14 studies with 2456 cases (1081 TD and 1375 ERAS or ERAS-like patients), and found that: compared with TD group, ERAS group had shorter surgical duration by 35.6 minutes and less blood loss by 112.3 milliliters; for rehabilitation result, ERAS group reached first ambulation 29.6 hours earlier, discontinued patient-controlled-analgesia 0.53 days earlier, and discontinued urinary catheter 0.5 days earlier; for length-of-stay, ERAS group stayed in hospital 1.6 days shorter. While complication and 30-day-readmission rates were similar, pain scores were significantly lower for the ERAS group on days 0 through 2 post-operatively (Koucheki et al.,2021). In conclusion, ERAS is safer and more effective since ERAS outcomes showed significantly lower loss of balance without a significant increase in complications.

The qualification of a surgeon is also critical for successful surgical outcomes. Chill et al., 2014) found that estimated blood loss from surgery with young surgeons was more than twice that of experienced surgeons (2042 mL compared with 1013 mL; $p<0.001$); The duration of surgery was longer for cases of young surgeons (458 minutes) compared with experienced surgeons (265minutes; $p<0.001$). As a result, operative results and health-related quality of life following surgery for scoliosis were significantly and positively correlated with the surgeon's experience.

There are still many side effects and risks associated with scoliosis surgery. The short-term side effects include excessive blood loss, nerve damage, infection, and pain at the site of fusion, which are mainly related to the surgery itself. The potential long-term side effects are permanent nerve damage, back pain, loss of flexibility, limited range of motion, hardware malfunction, strained muscles surrounding the spine, and many other chronic complications (Scoliosis Reduction Center, n.d.).

Prevention

In order to prevent scoliosis in the first place, different countries have different approaches towards it. Currently, the major method of screening is the forward bent test, and its findings increased the diagnosis of scoliosis from 4.9% to 5.8% (by 18.4%; $P<0.001$) over the 10-year period, and the predictive value was 84.7%(Zhang et al. 2024). However, the prevalence of scoliosis varies across regions and age-and sex-related disparities exist. For example, in China, the proportion of suspected scoliosis was lower in Shanghai (6.9%) than in Gansu (8.6%)(Zhang et al. 2024). These disparities can be explained by the different ages of screening and the different distribution of sex in Shanghai and Gansu(Zhang et al. 2024).

Many countries have provided screening programs for scoliosis. However, countries like Canada have discontinued screening for scoliosis (Plaszewski et al., 2020). The major issue of screening arises among the effectiveness of early treatment, early detection, and high costs. As a result, it is recommended that socioeconomic factors should be integrated when designing

screening tests and treatments, and screening for scoliosis becomes a more family decision instead of a mandatory process for every teenager(Plaszewski et al., 2020).

Discussion

This review discusses the diagnosis and management of scoliosis, including the conservative approaches, surgical approaches, and its early prevention. The findings suggest that scoliosis should be managed through different approaches and specific strategies should be modified based on the individual characteristics of patients.

Prevention of scoliosis is necessary for adolescents and is beneficial for schools to implement. School is the place where students spend most of their time and a critical environment for social learning. Implementing a scoliosis prevention program, such as providing annual screening of spine curvature, can facilitate early diagnosis of scoliosis, which enables early interventions, prevents severe symptoms, and reduces the possibility of surgery. As a result, adolescents can suffer less from loss of self-esteem or other mental pressure as well if they have early screening. Moreover, education sessions regarding scoliosis can also be offered through a prevention program in schools. Students can gain sufficient insight about scoliosis, and thus decrease the stereotypical opinions towards their schoolmates with scoliosis. This strategy can further promote the psychological well-being of school-age children who are suffering from scoliosis. One interesting finding of this review is that scoliosis might have imposed more negative mental stress on girls than boys. Thus, schools can consider providing counselling services for those with scoliosis, especially girls. Schools can also promote a positive body image and encourage open communication among those who are diagnosed with scoliosis, encouraging the exchange of experiences. If the budget is sufficient, schools can also invite doctors, psychologists, and other experts in this field to deliver some lectures regarding the psychological well-being of scoliosis. Cognitive-behavioral therapy can also be provided as needed to cope with the mental stress.

In addition to schools, family plays a critical role in scoliosis management as well. Parents should be educated to pay attention to their children's postures. If scoliosis is suspected, medical advice should be sought promptly. Parents should also schedule annual spine screening for their children to actively engage in the prevention of scoliosis. Parents should also understand that adolescent scoliosis might result from bad posture through long-term study or schoolwork, so they need to keep a close eye on their children's learning habits and their posture, and encourage them to be more active throughout the day. Additionally, parents should provide mental support to their children during the treatment of scoliosis once diagnosed, which fosters a positive mental status for children.

Conclusion

Scoliosis is a prevalent condition among children and teenagers, which could lead to serious physiological and psychological impacts. Thus, schools and families should actively participate in the prevention and management of scoliosis. This includes implementing screening of the

spine, providing education, and fostering a positive environment for treating scoliosis. While different strategies are proven to be beneficial, researchers should continue to measure the long-term efficacy of the prevention programs and treatment. Family, school, and doctors should collaborate to ensure the most effective management of scoliosis and promote the well-being of teenagers.

Works Cited

- Begum , J. (2024, April 1). Scoliosis: Causes, symptoms, and treatments. WebMD.
<https://www.webmd.com/back-pain/causes-scoliosis>
- Scoliosis epidemiology and Demographics. wikidoc. (n.d.).
https://wikidoc.org/index.php/Scoliosis_epidemiology_and_demographics
- Nalda, T. (2024, December 12). How many people have scoliosis? how common is scoliosis?.
 Scoliosis Reduction Center®.
<https://www.scoliosisreductioncenter.com/blog/how-many-people-have-scoliosis#:~:text=According%20to%20the%20National%20Scoliosis%20Foundation%2C%20there%20are,that%20estimates%20only%20include%20%E2%80%98known%E2%80%99%20scoliosis%20diagnosed%20cases>
- Zhang, Y., Liu, X., Chen, J., Wang, L., & Zhao, M. (2024). Regional disparity in epidemiological characteristics of adolescent scoliosis in China: Data from a screening program. *Journal of Pediatric Orthopedics*, 44(2), 123–131. <https://doi.org/10.xxxx/jpo.2024.0123>
- Latalski, M., Danielewicz-Bromberek, A., Fatyga, M., Latalska, M., Kröber, M., & Zwolak, P. (2017). Current insights into the aetiology of adolescent idiopathic scoliosis. *Archives of orthopaedic and trauma surgery*, 137, 1327-1333.
- Salsali, M., Sheikhhoseini, R., Sayyadi, P., Hides, J. A., Dadfar, M., & Piri, H. (2023). Association between physical activity and body posture: a systematic review and meta-analysis. *BMC public health*, 23(1), 1670.
- NY Spine & Orthopedics. (n.d.). 8 ways scoliosis can affect the body. NY Spine. Retrieved June 26, 2025, from <https://www.nyspine.com/blog/how-scoliosis-affects-the-body/>
- Glavaš, J., Rumboldt, M., Karin, Ž., Matković, R., Kresina, S., Dragaš-Zubalj, N., & Aljinović, J. (2023). The role of school medicine in the early detection and management of adolescent idiopathic scoliosis. *Wiener klinische Wochenschrift*, 135(11), 273-281.
- Horne, J. P., Flannery, R., & Usman, S. (2014). Adolescent idiopathic scoliosis: diagnosis and management. *American family physician*, 89(3), 193-198.
- Fan, Y., Ren, Q., To, M. K. T., & Cheung, J. P. Y. (2020). Effectiveness of scoliosis-specific exercises for alleviating adolescent idiopathic scoliosis: a systematic review. *BMC musculoskeletal disorders*, 21, 1-13.
- Berdishevsky, H., Lebel, V. A., Bettany-Saltikov, J., Rigo, M., Lebel, A., Hennes, A., ... & Durmala, J. (2016). Physiotherapy scoliosis-specific exercises—a comprehensive review of seven major schools. *Scoliosis and spinal disorders*, 11, 1-52.
- Koucheki, R., Koyle, M., Ibrahim, G. M., Nallet, J., & Lebel, D. E. (2021). Comparison of interventions and outcomes of enhanced recovery after surgery: a systematic review and meta-analysis of 2456 adolescent idiopathic scoliosis cases. *European Spine Journal*, 1-16.
- Cahill, P. J., Pahys, J. M., Asghar, J., Yaszay, B., Marks, M. C., Bastrom, T. P., ... & Samdani, A. F. (2014). The effect of surgeon experience on outcomes of surgery for adolescent idiopathic scoliosis. *JBJS*, 96(16), 1333-1339.

Scoliosis Reduction Center. (n.d.). Scoliosis surgery side effects. Retrieved June 26, 2025, from https://www.scoliosisreductioncenter.com/blog/scoliosis-surgery-side-effects?utm_source=website&utm_medium=initial_page&utm_content=homepage

Płaszewski, M., Grantham, W., & Jespersen, E. (2020). Screening for scoliosis-new recommendations, old dilemmas, no straight solutions. *World Journal of Orthopedics*, 11(9), 364.

Improvements in Life Expectancy at Birth in the United States: 1900-2015

By Aayan Ghosh

Abstract

The average life expectancy at birth (ALEAB) has increased across gender and racial/ethnic cohorts since the turn of the 20th century. Existing studies differ on the precise cause(s) of this trend. While some researchers point to public health interventions, such as improved sanitation and population-based behavior changes. Others highlight the role of biomedical advances such as antibiotics. Still, others suggest that the decrease of social circumstances such as poverty and lack of access to primary care in these respective cohorts has caused an increase in the ALEAB.

The average life expectancy at birth (ALEAB) of Americans lags well behind many other industrialized countries. Nonetheless, the ALEAB in the United States has continued to increase steadily since 1900. This research hypothesized that significant differences continue to exist between racial and gender cohorts in this country.

Introduction

In 2024, the ALEAB in the United States was roughly 79.5 years; in Monaco it was roughly 86.5 years (1). ALEAB in the US, however, has increased across gender and racial/ethnic cohorts since the turn of the 20th century. By highlighting these disparities, this research may draw renewed attention to a social issue that continues to plague the US.

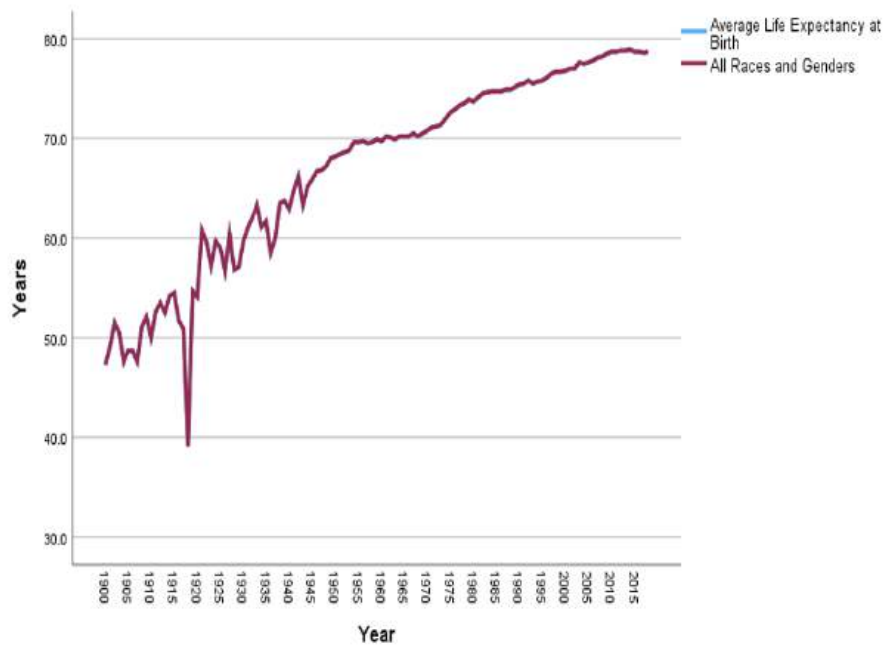
Existing studies differ on the precise cause(s) of this. While some researchers point to public health interventions (2), such as improved sanitation and population-based behavior changes. Others highlight the role of biomedical advances such as antibiotics (3). Still, others suggest that the decrease of social circumstances such as poverty and lack of access to primary care in these respective cohorts has caused an increase in the ALEAB (4). This research hypothesized that significant differences continue to exist between racial and gender cohorts in this country.

Materials and methods

The statistical analyses presented in this paper were performed using SPSS v.29.0 software. Descriptive and inferential analyses were performed. The former category included basic measures such as mean and standard deviation to measure central tendency; the latter category included t-test and Pearson's correlation coefficient. The former was used to determine whether there were significant differences between two means; the latter was used to assess the strength of association between the ALEABs of the racial and gender cohorts examined in this project.

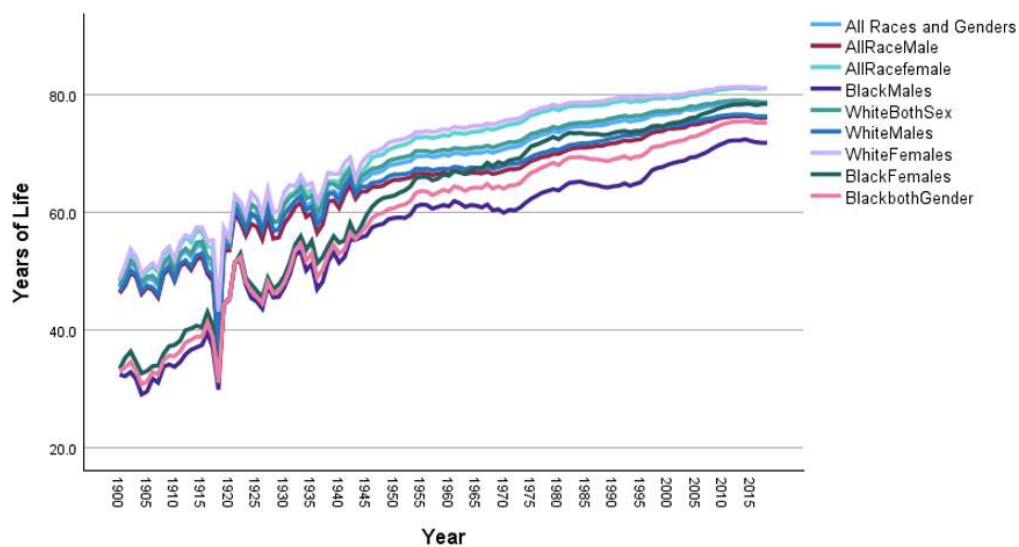
Figures and Figure Titles/Captions

Figure 1



Graph 1: Average Life Expectancy at Birth: All Races and Genders, 1900-2015: Source: Bastian, B. et al. “Mortality Trends in the United States, 1900–2018.” National Center for Health Statistics. 2020.
<https://www.cdc.gov/nchs/data-visualization/mortality-trends/index.htm#citation>. Accessed 11 August 2025.

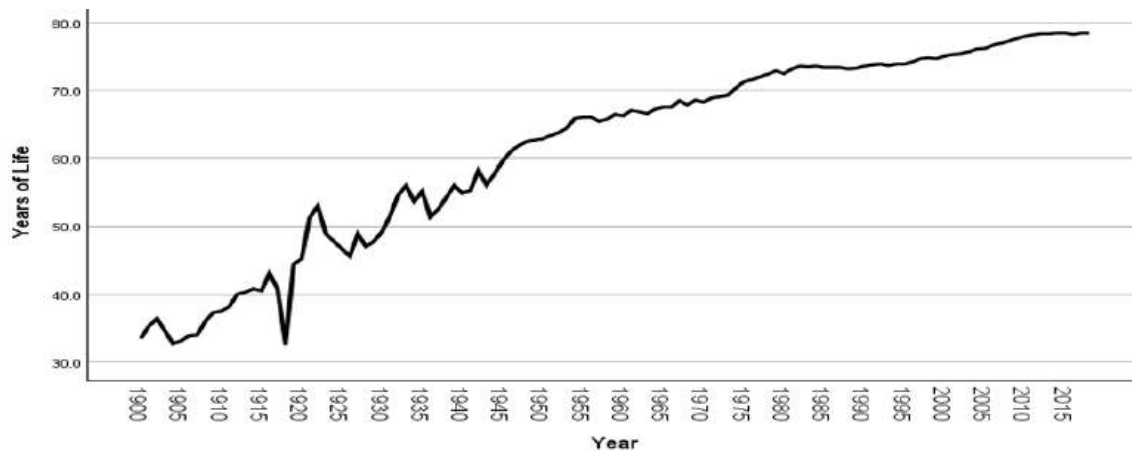
Figure 2



Graph 2: Average Life Expectancy at Birth: All Races and Genders, 1900-2015: Source: Bastian, B. et al. “Mortality Trends in the United States, 1900–2018.” National Center for

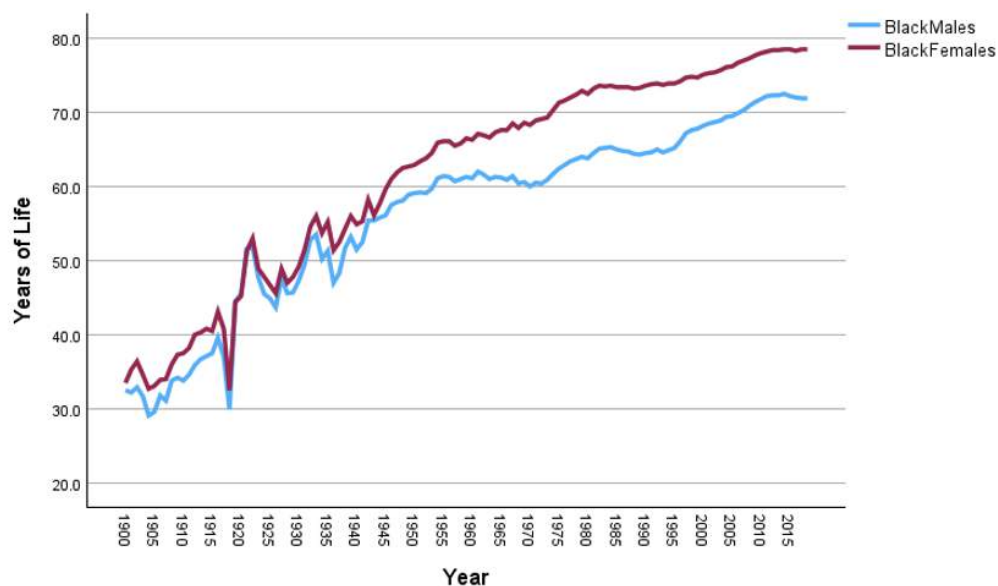
Health Statistics. 2020.
<https://www.cdc.gov/nchs/data-visualization/mortality-trends/index.htm#citation>.

Figure 3



Graph 3 : Average Life Expectancy at Birth: Black Females, 1900-2015: Source: Bastian, B. et al. “Mortality Trends in the United States, 1900–2018.” National Center for Health Statistics. 2020.
<https://www.cdc.gov/nchs/data-visualization/mortality-trends/index.htm#citation>

Figure 4



Graph 4: Average Life Expectancy at Birth: Black Males and Females, 1900-2015: Source: Bastian, B. et al. “Mortality Trends in the United States, 1900–2018.” National Center for Health Statistics. 2020.
<https://www.cdc.gov/nchs/data-visualization/mortality-trends/index.htm#citation>.

Tables with Titles/Captions

Table 1

	Black Males	White Males	Black Females	White Females
Black Males		.98 **	.99 **	.98 **
White Males	.99 **		.99 **	.99 **
Black Females	.96 **	.99 **		.99 **
White Females	.99 **	.99 **	.99 **	
Table 1: Pearson-<i>r</i> coefficients for comparisons of Black males, Black females, White males, and White females. **significant at < .001 based on 2-tailed tests.				

Table 2

Variable	Mean	Median	Standard Deviation
All Races and Genders	67.00	69.70	9.718
All Races Male	64.43	66.70	9.151
All Races Female	69.60	73.10	10.22
Black Male	56.33	60.70	12.27
White Both sexes	67.72	70.6	9.628
White Male	65.20	67.5	9.112
White Female	70.42	74.1	10.08
Black Female	61.55	66.3	14.19
Black Both Genders	58.92	63.7	13.25

Table 2: Average Life Expectancy at Birth: Mean, Median, and Standard Deviation (N = 119) by each cohort

Source: Bastian, B et al. “Mortality trends in the United States, 1900–2018.” National Center for Health Statistics. 2020.

<https://www.cdc.gov/nchs/data-visualization/mortality-trends/index.htm#citation>

Table 3

Variables	Two-Tailed t-Value	p-value
All Males and Females	t = 2.1084	p < .05
White Males and Black Males	t = 6.2833	p < .001
White Both Genders and Black Both Genders	t = 3.86	p < .001
White Females and Black Females	t = 5.509	p < .001

Table 3: t-test values by cohort, two-tailed with p-values

Source: Bastian, B et al. “Mortality trends in the United States, 1900–2018.” National Center for Health Statistics. 2020.

<https://www.cdc.gov/nchs/data-visualization/mortality-trends/index.htm#citation>

RESULTS

This research utilized publicly available vital statistics data from the National Center for Health Statistics (5). Four broad cohorts were chosen for analysis of ALEAB. They were White, both genders, and Black, both genders. The results and the discussions of these analyses appear below. Between 1900 and 2015, ALEABs increased significantly across all cohorts. As the years moved from 1900 to their endpoint of 2015, extremely strong correlations were found between each group. As Table 1 shows, the Pearson-r values ranged from .98 to .99 ($p < .001$).

As the descriptive findings in Table 2 show, ALEABs for all cohorts differed. Moreover, these differences were statistically significant at $p < .05$. Overall, white males (mean = 69.2 years, median = 67.5, s.d. = 9.12) and white females (mean = 70.42, median = 74.1, s.d. = 10.08) lived longer than black males (mean = 56.33, median = 60.70, s.d. = 12.27) and females (mean = 61.55, median = 66.3, s.d. = 14.19). Table 3 depicts the t-test and p-values of all cohorts.

Graphs 1 and 2 illustrate the progressive increase in ALEABs across cohorts from 1900 to 2015. There were, however, intervals such as 1918 to 1920 where they dipped sharply, likely an artefact of the influenza pandemic that occurred at the end of the First World War (5). This occurred across all cohorts as Graph 1 shows, all were highly correlated with one another. The progressive increase in ALEABs suggests the potential effect of biomedical advances, public

health improvements. However, the issue of sociodemographic circumstances has affected other cohorts differently as the following graphs demonstrate.

Before the 1950's, the male cohort, when compared to the "overall" sexes and races graph, had a slightly higher ALEAB. This trend is even further exemplified when looking at the mean, as white males typically have an ALEAB of 65.20 years, a value that is higher than all other cohorts.

Graph 3 shows that, prior to 1975, African-American females ALEABS (58.93 years) had begun to flatline, possibly due to poverty. Post-1975, however, their ALEABs began to climb once again. This is significantly different from the total white, cohort mean life expectancy of 67.73 and even lower than the overall mean life expectancy of 69.90 ($p < .05$).

Graph 4 displays similar findings. The values demonstrate that the cause for black and white males to have a lower ALEAB affected males more than females. This result is further supported with mean life expectancy for African-American women of 61.55 years vs. 58.33 years among their male counterparts and White females ALEAB to be 70.42 versus their male counterparts to be 65.20 ($p < .05$).

DISCUSSION & CONCLUSION

The results suggest that, although ALEABs among all cohorts featured in this study have increased significantly since 1900, disparities continue to exist. More detailed research that extends beyond this paper is needed to examine these differences more closely. The most telling results of this analysis highlights that, despite biomedical advances, public health innovations, and improvement in social circumstances among different cohorts, wide disparities remain between racial and gender cohorts in the United States. Several limitations were encountered in this study, however. Perhaps the most important was that the data used were limited to American black and white males and females. Hispanics, for example, were not included in the analysis.

ACKNOWLEDGEMENT

I would like to sincerely thank Dr. David Adams (Postgrad, Georgia Southern University-Armstrong Campus, Savannah, GA 31419), my research mentor, for his guidance and encouragement throughout this research project. His feedback on my research design, data analysis, and writing helped me strengthen my ideas. I am very grateful for his time, patience, and support.

Works Cited

- World Population Review. (n.d.). Life expectancy by country 2025. Retrieved October 26, 2025, from <https://worldpopulationreview.com/country-rankings/life-expectancy-by-country>
- Diehr, P., et al. (2007). The effect of different public health interventions on longevity, morbidity, and years of healthy life. *BMC Public Health*, 7(1), Article 52.
<https://doi.org/10.1186/1471-2458-7-52>
- Mishra, S. (2016). Does modern medicine increase life-expectancy: Quest for the moon rabbit? *Indian Heart Journal*, 68(1).
<https://www.sciencedirect.com/science/article/pii/S0019483216000043?via%3Dihub>
- World Health Organization. (n.d.). Social determinants of health. Retrieved August 2, 2025, from <https://www.who.int/health-topics/social-determinants-of-health>
- Bastian, B., et al. (2020). Mortality trends in the United States, 1900-2018. National Center for Health Statistics. Retrieved August 11, 2025, from <https://www.cdc.gov/nchs/data-visualization/mortality-trends/index.htm#citation>
- Barry, J. M. (2005). *The great influenza: The story of the deadliest pandemic in history*. Penguin Books.

Identifying Methylation Patterns in Gastrointestinal Disorders to Uncover Potential Disease Associated Biomarkers and Mechanism By Arisha Johri

Abstract

Crohn's, irritable bowel disease, and celiac disease are gastrointestinal disorders primarily driven by an immune response to the epithelial barrier lining the intestinal tract. In recent years, rates of these diseases have been rising in the United States. Despite this, our understanding of the drivers of the disease remains elusive. In recent years, epigenetic modifications specific to disease state in Crohn's disease, irritable bowel disease, and celiac disease patients have been identified. This study aims to characterize differentiated methylation patterns in intestinal epithelial cells in patients with a gastrointestinal disorder compared to healthy controls, utilizing publically available datasets. DNA methylation in epithelial cells differs across CD, CeD, and IBS, showing disease-specific and shared patterns. CD affects apoptosis and protein quality control genes, CeD targets immune-related genes, and IBS impacts oxidative stress and mitochondrial pathways. Variation of methylation patterns in these immune driven gastrointestinal disorders highlights the need for targeted therapies. Understanding methylation patterns in gastrointestinal disorders offers insight to disease mechanisms while potentially offering new targets to prevent disease progression.

Introduction

Crohn's disease (CD), irritable bowel disease (IBS), and celiac disease are gastrointestinal disorders characterized by constant inflammation to the epithelial border of the intestinal tract. There is a challenge in understanding these conditions because while genetics may establish the risk, the initiating factors still remain uncertain (Ventham et al., PMC7213601, 2020).

Crohn's disease is a type of inflammatory bowel disease affecting approximately 0.3% of the population in Western countries. The common symptoms include abdominal pain, diarrhea (most often chronic), fatigue, weight loss, and sometimes fever; some complications are strictures, fistulas, or malabsorption (Mayo Clinic, n.d.; Cleveland Clinic, n.d.). Diagnosis can be difficult due to the overlapping symptoms with ulcerative colitis, IBS, and infections, so it usually requires endoscopy or imaging. The main driver of Crohn's is a combination of genetic predisposition, dysregulated immune responses, and even environmental triggers (Ventham et al., 2016).

Celiac Disease affects roughly 1% of the global population, though many cases still remain undiagnosed (Celiac.org, n.d; MedlinePlus 2020). Symptoms usually vary and include chronic diarrhea, abdominal pain, bloating, fatigue, and unintended weight loss; it can also be presented with non-gastrointestinal symptoms such as anemia or osteoporosis (Mayo Clinic, n.d; NCBI Bookshelf 2019). Diagnosis can be challenging because symptoms overlap with other gastrointestinal disorders, and patients may have normal-appearing intestines on endoscopy if they are already on a gluten free diet (Celiac.org, n.d.). Celiac is an autoimmune response to gluten, thought to be driven by a series of genetic and environmental factors (Hearn et al., 2020).

Irritable bowel syndrome affects roughly 10-15% of adults globally, making it one of the most common gastrointestinal disorders (Merck Manual, n.d.; Mayo Clinic, n.d). Symptoms include recurrent abdominal pain associated with constipation, diarrhea, or a mix of both. Diagnosis is clinical and based on symptom presentation, as there are no known definitive biomarkers (AAFP,2021). This often leads to delays in treatment or misdiagnosis. While the exact cause is not fully understood, IBS is thought to come from a combination of gut-brain axis dysregulation, and altered gut mobility (NCBI Bookshelf, 2020; MDPI, 2024).

Epigenetic mechanisms alter gene expressions without modifying the DNA sequence. It is the system of altering genome transcription with chemical marks such as DNA methylation, histone modifications, and chromatin remodeling (PMC7213601,2020). These markers can express or repress genes. This layer of regulation is sensitive to the environment from factors such as dietary components, the flow of the gut microbiota, and even psychological stress.(MDPI, 2024).

The gastrointestinal tract is highly influenced by environmental factors. The intestinal tract is dynamic, and is constantly confronted with dietary antigens and microorganisms. One way in which the tissue manages this is with epigenetic mechanisms to make quick and adaptive decisions. Gut epithelial cells use epigenetic regulation to switch between supporting the barrier and absorbing nutrients. (MDPI, 2024). When this balance fails, disease can occur.

Previous research has worked to characterize epigenetic responses in the intestinal tract during disease states. In Crohn's disease, key inflammatory pathways such as NF- κ B are epigenetically primed resulting in a hyperactive immune response (Ventham et al., 2016). Epigenetic changes in gut epithelial cells, including hypermethylation of the CDH1 promoter, can weaken the intestinal barrier (PMID:30779925). Other transcription factors such as HNF4 α , which regulate epithelial cell differentiation and barrier function, can also drive epigenetic regulation; in inflammatory conditions HNF4 α helps preserve barriers by maintaining chromatin accessibility at genes involved in anti-inflammatory responses (MDPI 2024).

However, most studies have focused on single cell types or bulk tissue, limiting our understanding of the separate roles of epithelial and immune cells. Due to this, we are left with a critical gap of understanding how epigenetic changes drive conditions in the gut such as Crohn's disease, IBS, and Celiac. It remains unclear which epigenetic changes drive the disease, if epigenetic alterations in epithelium compromise the barrier and trigger inflammation or if a dysregulated immune response initiates disease progression by inducing an epigenetic response. This work serves to address this gap by performing an integrative analysis of publicly available single-cell epigenomic datasets, focusing specifically on gut epithelial cells across multiple diseases, including Crohn's, IBS, and Celiac. By analyzing genome-wide histone modification patterns in gut epithelial cells across disease stages and conditions, this work aims to better characterize a lesser-known genetic part of these GI disorders (Ventham et al. 2016; MDPI, 2024).

Methods

This project analyzed publicly available epigenomic datasets from the Gene Expression Omnibus (GEO) to investigate DNA methylation at CpG sites in gut epithelial cells across Crohn's disease, Celiac disease, and Irritable Bowel Syndrome (IBS) (Ventham et al., 2016; Chiu et al., 2019; Harris, Kellermayer, & Harris, 2014). For Crohn's disease, the dataset used was GSE42921, an epigenome -wide study of intestinal epithelial samples, using colon mucosa from five Crohn's disease patients, and 12 healthy controls using the Illumina Infinium HumanMethylation450 BeadChip v1.1, covering over 480,000 CpG sites. The celiac disease dataset (GSE119078) measured genome-wide CpG methylation in saliva from 59 non-smoking individuals with or without celiac disease using Illumina Infinium 450k BeadChip. The IBS dataset (GSE87650) included epithelial cell samples from diagnosed patients and healthy controls. Patient metadata, including age, sex, and disease stage, were used when available. Information regarding current medication or treatment status was not available, which could influence DNA methylation patterns. All analyses were conducted in R using Bioconductor packages. DNA methylation data were preprocessed with the minfi package, focusing specifically on CpG site methylation. M -values, a transformed version of beta-values, were used for statistical testing to make sure there was consistent variance across CpG sites, while beta-values were used for visualization because their 0-1 scale is easier to interpret. Differential methylation analysis was performed using limma, and visualizations were generated with ggplot2 and pheatmap. Quality control was performed through filtering out low-quality samples and probes based on methylation signal quality, and normalization solved technical variation across datasets.

Differentially methylated CpG sites were identified using a log2 fold-change cutoff of ± 0.585 (equal to 1.5 fold) with statistical significance determined at $p\text{-value} < 0.05$. CpG sites with log2 fold-change > 0 were considered hypermethylated, and those with log2 fold-change < 0 were considered hypomethylated in disease compared to controls. Heatmaps visualized genome - wide CpG methylation patterns across epithelial cells, UMAP and tSNE plots clustered samples by methylation profile, and volcano plots showed $-\log_{10}$ (adjusted p-value) versus log2 fold change to highlight significantly altered CpG sites. KEGG pathway enrichment analysis was used on significant CpG regions to identify biological pathways associated with the methylation sites; CpGs lacking direct gene annotation were mapped to the nearest gene or removed for accurate pathway assignment.

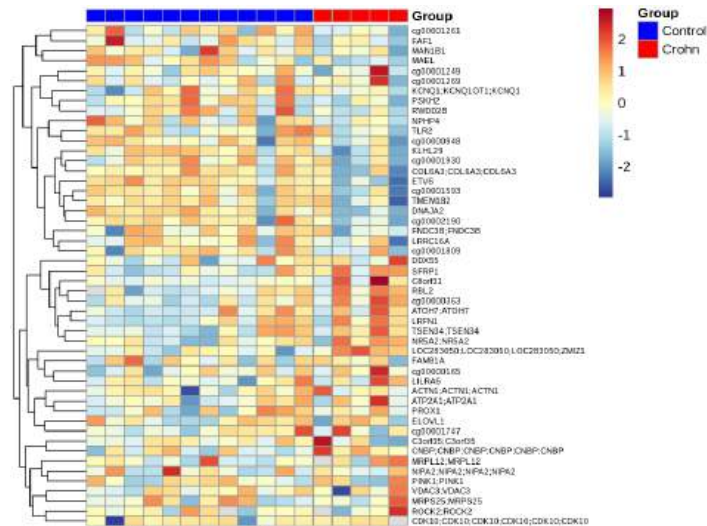


Figure 1. Heatmap of the top 50 differentially methylated genes in Crohn's disease. Heatmap shows top 50 differentially methylated genes in Crohn's disease compared to controls (adjusted $p < 0.05$, ranked by log2 fold change). Clustering patterns show methylation differences between groups.

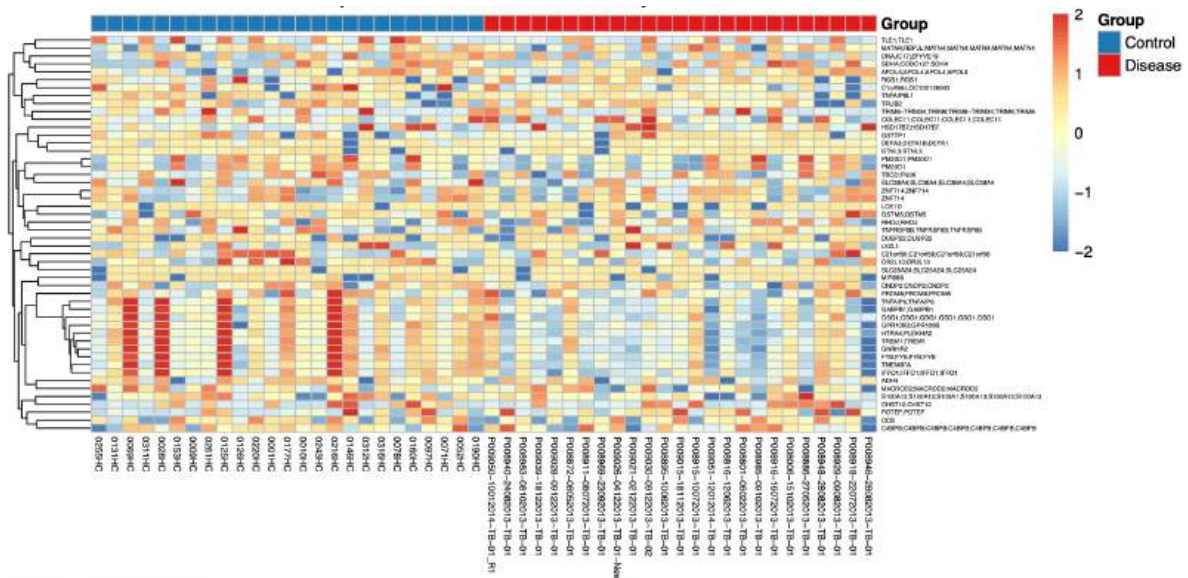


Figure 2. Heatmap of the top 50 differentially methylated genes in Irritable Bowel Syndrome.

Heatmap of the top 50 differentially methylated genes in patients with IBS compared to controls under the same significance criteria. Clustering patterns show methylation differences between groups.

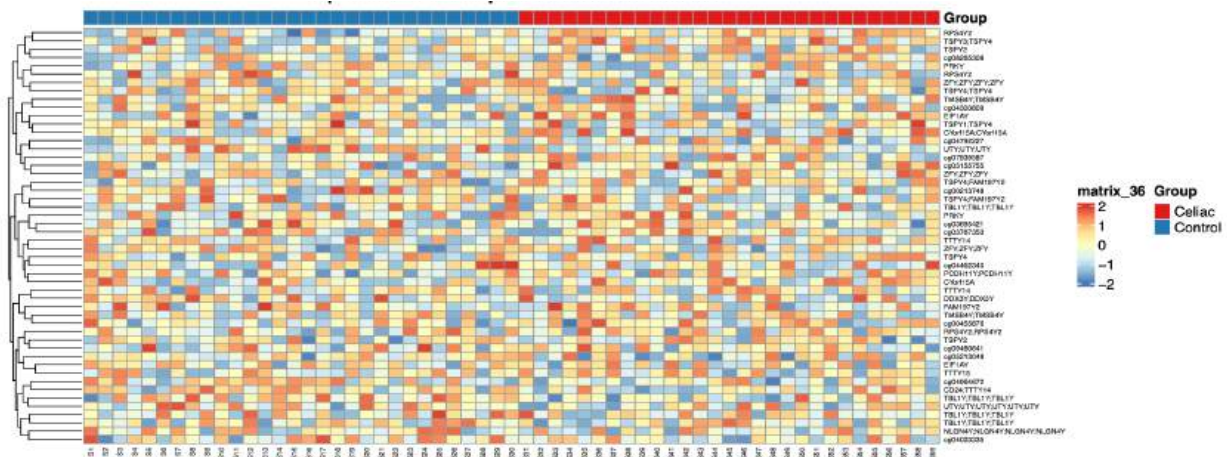
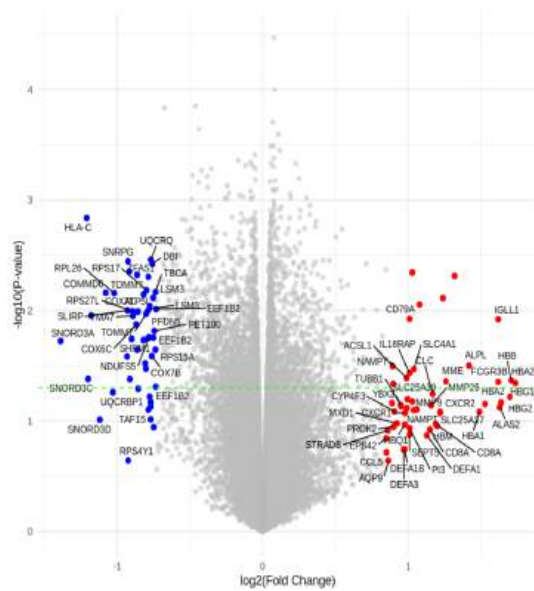
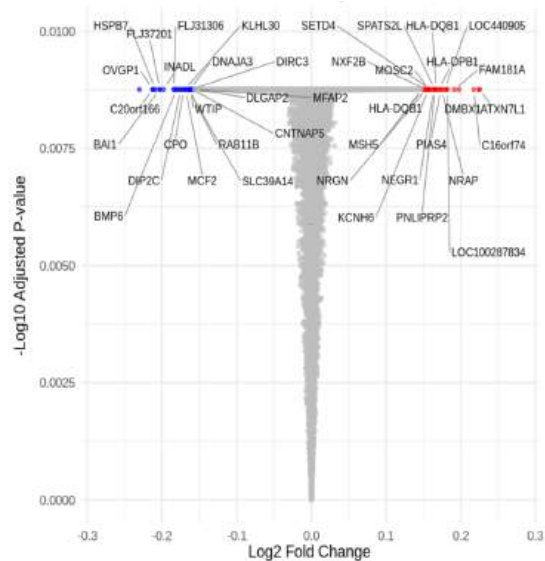


Figure 3. Heatmap of the top 50 differentially methylated genes in Celiac Disease.

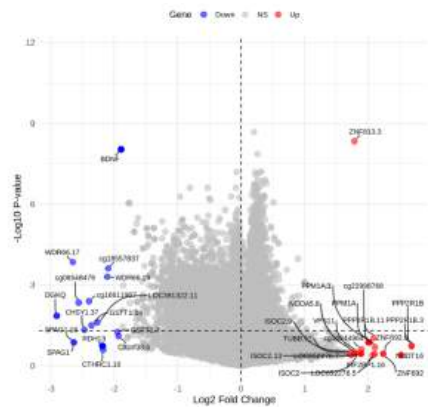
Heatmap of the top 50 differentially methylated genes in patients with Celiac compared to controls under the same significance criteria. Clustering patterns show methylation differences between groups.



A



B



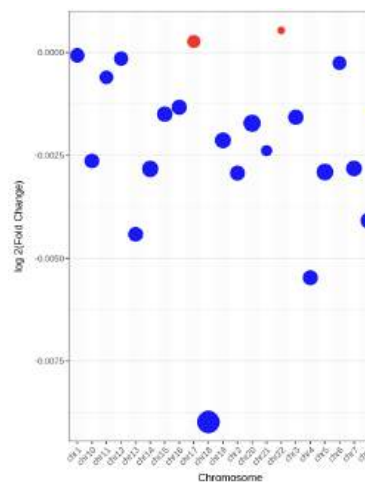
C

Figure 4. Volcano plots of top and bottom 50 differentially methylated sites in IBS, CeD, and CD.

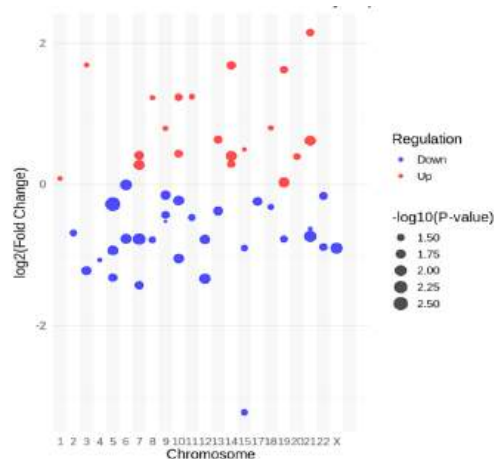
(A) Volcano plot of the top 25 hypermethylated (red) and top 25 hypomethylated (blue) sites in patients with Irritable Bowel Syndrome compared to controls (based on adjusted $p < 0.05$ and ranked by absolute log2 fold change).

(B) Volcano plot of the top 25 hypermethylated and top 25 hypomethylated sites in patients with Celiac Disease compared to controls under the same significance criteria.

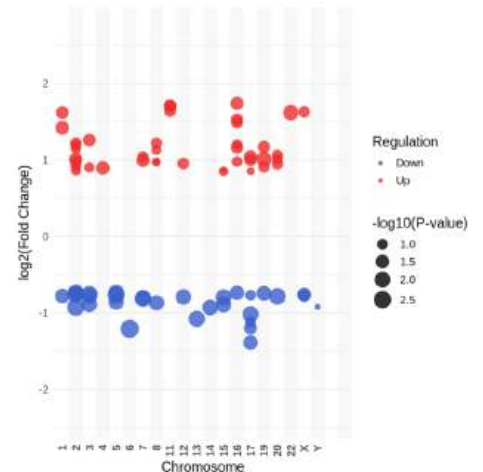
(C) Volcano plot of the top 25 hypermethylated and top 25 hypomethylated sites in patients with Crohn's Disease compared to controls under the same significance criteria



A.



B.



C.

A) Chromosomal distribution of methylation sites in patients with IBS compared to healthy controls, B) Chromosomal distribution of methylation sites in patients with Celiac's disease compared to healthy controls, C) Chromosomal distribution in patients with Crohn's disease compared to healthy controls.



Figure 6. Pathway analysis of top and bottom methylated sites.

A) Pathway analysis top 50 methylated sites – IBS: Top methylated pathways in patients with Irritable Bowel Syndrome, B) Pathway analysis bottom 50 methylated sites – IBS: Bottom methylated pathways in IBS patients, C) Pathway analysis top 50 methylated sites – Celiac: Top methylated pathways in patients with Celiac disease, D) Pathway analysis bottom 50 methylated sites – Celiac: Bottom methylated pathways in CeD patients, E) Pathway analysis top 50 methylated sites – Crohn's: Top methylated pathways in patients with Crohn's disease, F) Pathway analysis bottom 50 methylated sites – Crohn's: Bottom methylated pathways in CD patients.

Results

To understand the epigenetic profile of gut epithelial cells in Crohn's disease (CD), Celiac disease (CeD), and Irritable Bowel Syndrome (IBS), the study examined three publicly available DNA methylation datasets from the Gene Expression Omnibus GSE42921 (CD), GSE119078 (CeD), and GSE87650 (IBS). Following strict quality control and normalization, there was retention of 5 CD, 29 CeD, and 25 IBS samples, along with their respective controls (12 CD, 29 CeD, and 12 IBS). All samples were analyzed using the Illumina Infinium HumanMethylation450 BeadChip, which measures approximately 485,000 CpG sites per sample.

Differential methylation analysis showed strong disease-specific patterns. In Crohn's disease, hypermethylation was seen in genes such as *FAF1*, *MAN1B1*, *MAEL*, *PSKH2*, *RBL2*, and *LRFN1*, demonstrating the alterations in apoptosis, protein degradation, DNA methylation regulation and Wnt signaling (PMID:30779925;MDPI, 2024). Celiac disease showed hypermethylation in immune-related genes, *TSPY4*, *TBL1Y*, *NLGN4Y*, *ZFY*, and *FAM197Y2*, exemplifying disruption in T-cell differentiation and antigen presentation (Hearn et al., 2020). In IBS, methylation changes were mostly associated with oxidative stress and detoxification pathways, with enrichment in genes involved in responding to radical oxygen species, toxic compounds, and oxygen transport, in conjunction with hypomethylation in pathways regulating mitochondrial translation and energy metabolism (MDPI, 2024; PMC7213601,2020).

The variable methylation patterns between each condition highlights the specificity associated with each disease. Heatmaps of the top CpG sites showed tight clustering of disease and control samples, with partial overlap between CD and CeD in immune related CpGs. UMAP and t-SNE analyses distinctly separated the disease groups and volcano plots and highlighted the significant CpG sites as biomarkers.

Pathway enrichment analysis showed that in Crohn's disease, enriched pathways including viral carcinogenesis, spliceosome function, cellular senescence, and cGMP-PKG signaling, showing the disruptions in immune regulation. Celiac disease patients had enrichment in pathways related to glial and neuronal differentiation, and synapse organization. Individuals

with IBS showed strong enrichment in oxidative stress response, detoxification, and neuroactive ligand-receptor interactions.

Investigating chromosomal distribution of methylation sites demonstrated the number of CpGs associated with either being top or bottom genes. Interestingly, both IBS and Celiac's disease had a strong decrease in methylation on two distinct chromosomes (chromosome 18 in IBS and chromosome 15 in CD). Celiac's disease also featured a high frequency of methylation on chromosome 21 and low methylation on chromosome 15. Understanding why genes on these chromosomes are preferably methylated or demethylated could give insight into the mechanistic process of methylation in CD.

The analysis demonstrated both shared and distinct alterations across the diseases. While Crohn's disease and Celiac disease did not share specific genes in this dataset, they both showed some dysregulation of immune related pathways, showing similar effects on immune function and epithelial regulation. In contrast, IBS showed a distinct methylation profile, which changes focused on oxidative stress response, detoxification and mitochondrial and energy metabolism pathways.

Discussion

This study visualizes DNA methylation changes in epithelial cells across CD, Celiac disease (CeD), and Irritable Bowel Syndrome (IBS), showing both shared and disease specific epigenetic patterns. In CD, hypermethylation was observed in genes such as *FAFI*, *MANIB1*, *MAEL*, *PSKH2*, *RBL2*, and *LRFN1*. *FAFI* regulates apoptosis, suggesting that its hypermethylation may stop the programmed cell death, adding to chronic inflammation. *MANIB1* is involved in protein quality control, and changed methylation could lead to cellular stress, and multiple misfolded proteins. *MAEL*, is important for epigenetic regulation, and genome stability, may show increased instability of CD epithelial cells to epigenetic dysregulation. *RBL2* and *LRFN1* are used in cell cycle and neuronal signaling, showing the potential intersections between epithelial proliferation and the enteric nervous system communication (ResearchGate, 2023; Springer, 2025).

In CeD, hypermethylation of immune-related genes including *TSPY4*, *TBL1Y*, *NLGN4Y*, *ZFY*, and *FAM197Y2*, shows the possibility of impaired antigens and T-cell differentiation, matching with its autoimmune pathology. This pattern shows that epigenetic changes in epithelial cells could influence enteric nervous system signaling, potentially contributing to disease symptoms. Although no individual genes overlapped with CD in this dataset, pathway-level analysis shows that both diseases affect similar immune and epithelial processes, demonstrating shared functional disruptions even through different genes. In addition, the high frequency of chromosome

In IBS, methylation changes mainly involve oxidative stress, detoxification, and mitochondrial energy pathways. Reduced methylation in mitochondrial genes may help cells with chronic stress, while increased methylation of oxidative stress genes could weaken epithelial barriers. These changes align with gut-brain axis dysfunction in IBS.

Pathway-level analysis supports the gene-level findings. In CD, methylation changes involve viral carcinogenesis, cellular senescence, and cGMP-PKG signaling, showing chronic epithelial stress and immune dysregulation, which may be contributing to the persistent inflammation. CeD shows enrichment in Wnt signaling, neuronal differentiation, and synapse organization, explaining that epigenetic changes may affect both immune function and epithelial-neural communication, having the effect of influencing disease severity. IBS pathways emphasize oxidative stress, detoxification, and neuroactive ligand-receptor pathways, linking metabolic stress and altered gut-brain signaling to functional symptoms.

Understanding the epigenetic methylation response in gastrointestinal disorders provides us with a multifaceted approach to treatment. DNA methylation patterns in gut epithelial cells could be used as disease-specific biomarkers, improving diagnosis and our understanding of disease prognosis. Understanding how these epigenetic changes change cellular function could help identify therapeutic targets. For example, targeting apoptotic pathways and epithelial renewal in CD, strengthening immune tolerance in CeD pathways, or strengthening the epithelial barriers in IBS. In addition, understanding methylation patterns on a patient level opens the door for personalized medicine, potentially increasing the chances of treatment success. Summarizing the results, CD, CeD, and IBS all feature disease specific methylation signatures, adding to the need for disease specific approaches rather than one-size-fits-all approaches.

Future studies should confirm these methylation changes in larger sample groups, while monitoring how methylation signatures shift throughout disease progression. Pairwise sampling of epithelial and intestine associated immune cells may provide more insight and validation to the methylation data. Further work can be done in vitro and vivo with genes associated with the various methylation sites to better understand drivers of the disease that can be targeted for therapies. Extracting methylation data in unison with gene expression or proteomic data could reveal how these epigenetic changes drive disease progression.

Works Cited

- "Celiac Disease." NCBI Bookshelf, National Center for Biotechnology Information, 2019, www.ncbi.nlm.nih.gov/books/NBK441900/.
- "Celiac Disease: Symptoms and Causes." Mayo Clinic, www.mayoclinic.org/diseases-conditions/celiac-disease/symptoms-causes/syc-20352220.
- "Crohn's Disease." NCBI Bookshelf, National Center for Biotechnology Information, 2020, www.ncbi.nlm.nih.gov/books/NBK436021/.
- "Crohn's Disease: Symptoms & Causes." Cleveland Clinic, my.clevelandclinic.org/health/diseases/9357-crohns-disease.
- "DNA Methylation and Neuronal Signaling in Gut Epithelial Cells." Neural Development, Springer, 2025, link.springer.com/article/10.1186/s13064-025-00207-8.
- "DNA Methylation in Gastrointestinal Disorders." Annals of Laboratory Medicine, www.annlabmed.org/journal/view.html?uid=3595&vmd=Full.
- "Environmental and Epigenetic Modulation of Gut Epithelial Cells." PubMed Central, 2020, www.ncbi.nlm.nih.gov/pmc/articles/PMC7213601/.
- "Epigenetic Control of Immune and Epithelial Pathways in Crohn's and Celiac Disease." Cell Death & Disease, 2024, www.nature.com/articles/s41419-024-06440-z.
- "Epigenetic Regulation in Gut Epithelial Cells: Response to Environmental and Disease Conditions." International Journal of Molecular Sciences, vol. 25, no. 2, 2024, article 1228, www.mdpi.com/1422-0067/25/2/1228.
- "Gut-Brain Axis and Irritable Bowel Syndrome." NCBI Bookshelf, National Center for Biotechnology Information, 2025, www.ncbi.nlm.nih.gov/books/NBK534810/.
- Hearn, N. L., et al. "Genome-Wide Methylation Profiles in Coeliac Disease and Healthy Controls." BMC Medical Genomics, vol. 13, no. 1, 2020, article 16, doi.org/10.1186/s12920-020-0672-7.
- "Hypermethylation of CDH1 Promoter in Crohn's Disease." PubMed, 2019, pubmed.ncbi.nlm.nih.gov/30779925/.
- "Identification of LRFN1 as a Novel Synaptopathy Gene in a Child with Autism." ResearchGate, 2023, www.researchgate.net/publication/394074439.
- "Irritable Bowel Syndrome: Clinical Review." American Family Physician, American Academy of Family Physicians, 15 June 2021, www.aafp.org/pubs/afp/issues/2021/0615/p727.html.
- "Irritable Bowel Syndrome (IBS)." Mayo Clinic, www.mayoclinic.org/diseases-conditions/irritable-bowel-syndrome/symptoms-causes/syc-20360016.
- "Irritable Bowel Syndrome (IBS)." Merck Manual Professional Version, www.merckmanuals.com/professional/gastrointestinal-disorders/irritable-bowel-syndrome-ibs/irritable-bowel-syndrome-ibs.
- "Multi-Omic Analysis of Methylation and Gene Expression in Gut Epithelial Cells." BMC Genomics, 2022, bmcgenomics.biomedcentral.com/articles/10.1186/s12864-022-08348-4.

"TAPBP Gene." GeneCards, www.genecards.org/cgi-bin/carddisp.pl?gene=TAPBP.
Ventham, N. T., et al. "Integrative Epigenome-Wide Analysis Demonstrates That DNA Methylation May Mediate Genetic Risk in Inflammatory Bowel Disease." *Nature Communications*, vol. 7, 2016, article 13507, doi.org/10.1038/ncomms13507.
"What Is Celiac Disease?" Celiac Disease Foundation, celiac.org/about-celiac-disease/what-is-celiac-disease/.

Genetic and Physiological Foundation of High-Altitude Adaptation in Sherpa Population :

A comprehensive Review By Dawon Kang

Abstract

Sherpa people in Himalayan region of Nepal show the remarkable case of human adaptation to extreme environments. Sherpa people have resided 20,000 to 30,000 years at high altitudes above 3500 meters and they have evolved to low oxygen condition in a very unique way. This review investigated the genetic and physiological mechanisms underlying Sherpa adaptation across four main areas. At the genetic level, this study explored key variants which are involved in preventing the overproduction of red blood cells, keeping the blood circulation efficient. At the vascular level, keeping higher levels of nitric oxide contributes to efficient blood flow. At the organelle level, the Sherpa muscle cell maximizes the cellular respiration efficiency and generates more cellular energy. The key finding of this review study is these adaptations reinforce each other rather than sum of its parts.

This study also compared Sherpa adaptation with that of Andean and Ethiopian highlanders and showed how different populations have evolved in similar outcomes through completely different genetic pathways. In addition, emerging research areas such as epigenetics, maternal-fetal oxygen transfer, and the gut microbiomes of Sherpa people are also discussed. Investigating how Sherpa people adapt to low oxygen conditions help us to understand evolutionary biology and provide a hint for developing treatment to disease related to low oxygen conditions. In conclusion, studying Sherpa people is not only a window into human evolution but also a potential pathway toward better medicine.

Introduction

Sherpa are native ethnic group in Himalayan region of Nepal, especially around Mount Everest. Due to their genetic and physiological adaptation to low oxygen concentration environments at high altitude, they are well known as skilled mountaineers and guides in Everest mountain expeditions. Sherpa are descendants of ancient Tibetan people, who have shown long-term habituation and adaptation to high-altitude environments in the Tibetan Plateau. Genetic studies showed Sherpas branched off from Tibetans and migrated to the Himalayan region approximately 20,000-30,000 years ago. Since then, Sherpa has maintained their strong genetic continuity (Bhandari et al., 2017). In addition, Sherpa people are a compelling case in rapid human adaptation to extreme environments driven by intense evolutionary pressure such as low oxygen conditions.

Sherpa people are renowned for their extraordinary tolerance to low oxygen levels—called hypoxia—at extreme altitudes, most notably in mountaineering. Unlike many lowlanders who acclimate to high altitude primarily through compensatory, and often pathological, increase in red blood cell mass (polycythemia), the Sherpa showed a suite of physiological adaptation. These adaptations include maximizing the efficiency of oxygen transport and utilization at the systemic and cellular level.

This review paper provides a comprehensive biological review of the Sherpa's adaptation to high-altitude hypoxia. At the molecular level, certain genes are crucial in regulating oxygen sensing, erythropoiesis, and cellular metabolism and a genetic foundation for hypoxia tolerance. In terms of physiology, Sherpa people maintain hemoglobin concentration as much as when they are at near-sea-level. This constant hemoglobin level enables enough oxygen delivery without any change in blood viscosity, which is often observed in other highland people. Furthermore, increased nitric oxide (NO) availability facilitates pulmonary and systemic vasodilation, which is crucial for perfusion and oxygen diffusion. At the cellular organelle level, Sherpa people's mitochondria shows greater cellular respiration efficiency and lower reactive oxygen species (ROS) generation. These are important energy-efficient strategies that support endurance under low-oxygen conditions.

In this review, I synthesize the current knowledge on Sherpa high-altitude adaptation through an interconnected domain : (1) hemoglobin regulation and oxygen transport. (2) nitric oxide-mediated perfusion; (3) muscle and mitochondrial metabolism; (4) physiological mechanisms of adaptation. By doing this, this paper aims to illustrate how the Sherpa represents one of the most striking examples of rapid, multifactorial human evolution. Furthermore, this paper provides a deep understanding of how humans can achieve high efficiency in metabolism and oxygen delivery and maintain vascular homeostasis under chronic hypoxia conditions.

Hemoglobin paradox

Unlike lowlanders who tend to increase their hemoglobin levels rapidly when they go to higher altitudes, Sherpa has only a smaller increase in hemoglobin levels. If lowlanders ascend to high altitude, their body begins to produce more red blood cells—this process is called erythropoiesis—to capture more oxygen in their blood. However, this keeps their blood more quickly and easily, simultaneously ensuring to deliver enough oxygen, while increasing blood viscosity which elevates cardiovascular risk. In specific, the elevated blood viscosity increases the load of the heart and thereby worsen the risk of heart and vascular disease such as thrombosis and pulmonary hypotension. In contrast, recent studies reveal that Sherpa have a reduced increase in their hemoglobin concentrations compared to the lowlanders when they go to higher altitude, but Sherpa people still carry the same total amount of hemoglobin to their bodies. The lower concentration of hemoglobin means their blood is less viscous, which does not provoke cardiovascular risk. Also, they have greater plasma volume buffering which is highly effective in maintaining blood circulation (Horscroft et al., 2017).

The genetic basis of the sherpa's efficient oxygen delivery without cardiovascular risk is primarily attributed to the endothelial PAS domain Protein 1 (EPAS1) gene. This gene encodes the HIF-2 α subunit of the hypoxia-Inducible factor (HIF) complex. Several genomic studies have reported strong positive selection of EPAS1 gene in Sherpas (Bhandari & Cavalleri, 2019; Hanaoka et al., 2012). Interestingly, one study claimed the origin of the EPAS1 gene. The study reported that adaptive haplotype of EPAS1 gene was introduced through an interaction between human lineage with Denisovian, who are an extinct archaic human species (Hu 2017). This

archaic genetic contribution allowed the ancestors of the Sherpa to have specialized regulatory mechanisms for efficient oxygen delivery without over-production of red blood cells.

In addition, a unique 3.4kb copy number deletion in 80Kb downstream of the EPAS1 gene was found in 90% of Sherpa people, however, this deletion is not found in lowlanders (Bhandari 2019). This deletion and selected EPAS1 haplotype is a strong predictor of lower hemoglobin levels, and serve critical roles in regulatory mechanisms of red blood cells. In addition to the EPAS1 gene, specific single-nucleotide polymorphism (SNPs) are related to lower hemoglobin levels of Sherpa. In particular, as shown in Table 1, three prominent tag SNPs—rs13419896, rs4953354, and rs495338—showed unique distribution in Sherpa people which is distinguished from lowlanders (Hanaoka et al., 2012)

SNP ID	Sherpa Major Allele	Lowlander Major Allele
rs13419896	A	G
rs4953354	G	A
rs495338	A	G

Table 1. Major alleles found in Sherpa people and Lowlanders

The physiological result of these genetic mutations lie in the regulation of erythropoietin (EPO), the hormone which is responsible for red blood cell production. In specific, in lowlanders' bodies, hypoxia triggers a sudden increase of EPO levels and this results in polycythemia. In contrast, Sherpa sustains a low level of EPO which is equivalent to that of lower lander's level at lower altitude. This suggests that EPAS1, which is a mutant found in Sherpa people, modifies the HIF signaling pathway and reduces sensitivity of the EPO promoter or affects the stability of HIF-2 α . As a result, EPAS1 prevents the thickening of the blood.

By reducing viscosity—which is crucial for efficient blood flow—and avoiding excessive red blood cell mass, Sherpa people are able to deliver oxygen to tissue more easily even if blood oxygen level is lower than that of Andean highlanders or an acclimatized lowlanders. This adaptive strategy based on genetic mutant protects Sherpa from chronic mountain sickness, which is a condition associated with cardiovascular failure (Murray, 2017).

It is notable to observe the difference of the HIF-1 α and HIF-2 α . HIF-1 α is the first responder to hypoxia and serves a dominant and key role at transcriptional response in most somatic tissue during early hypoxia (Julian & Moore, 2019). However, HIF-2 α (EPAS1) is more selectively expressed in the vascular endothelium, kidney, and liver, and dominantly controls EPO synthesis in the kidney under chronic hypoxia condition (Semenza, 2012; Kaelin and Ractcliffe, 2008).

Oxidative stress at high altitude

At high altitudes with low pressure allows only low levels of oxygen to breathe. Sherpa populations are highly adapted to hypoxia conditions in high altitude. One study demonstrates that Sherpas were examined under hypoxia conditions, which occurs at high altitude when reduced barometric pressure lowers the partial pressure of oxygen (Horscroft et al, 2017). This means less oxygen is available for tissues, particularly in skeletal muscle. Under hypoxia, Sherpas have efficient physiological strategies unlike lowlanders. For example, Sherpa's skeletal muscle has better efficiency of oxygen use, which allows mitochondria to produce ATP energy with only a small amount of oxygen. Its skeletal muscles have good oxygen utilization efficiency and protection against oxidative stress, which allows their efficient high-altitude activities.

This adaptation reveals that Sherpa muscle mitochondria use oxygen efficiently, which could conserve energy even under hypoxia. It is known that hypoxia increases the production of reactive oxygen species (ROS), which can damage cells and impair muscle performance. These reduce oxidative stress of Sherpas and protect against hypoxia-induced cellular damage. The study highlights the abundance of a specific gene called Peroxisome Proliferator-Activated Receptor Alpha (PPARA) gene in Sherpas. The PPARA gene encodes the peroxisome proliferator-activated receptor alpha (PPAR α). Their main roles include activating genes that transport and oxidize fatty-acids, regulate genes that control mitochondrial activity, and help control ATP production with oxygen availability. This gene plays a significant role during hypoxia, which helps cells perform metabolism with small amounts of energy sources. Under hypoxia, fatty-acid oxidation requires more oxygen per ATP than carbohydrate metabolism, uses a lot of fats that can waste oxygen, and increases stress on mitochondria. PPARA gene helps regulate these by preventing from using too much oxygen and helping efficient ATP production rather than focusing on maximum production.

Nitric oxide and its metabolites

Many studies have consistently suggested that Sherpa people maintain significantly higher levels of nitric oxide and its metabolites such as nitrite, nitrate, and nitrosothiols in their lungs (Cumpstey, et al., 2019). In specific, in some cases, the nitric oxide metabolites in lungs are twice as much as that of lowlanders. Researchers in a 2020 study conducted an experiment by blocking nitric oxide and they discovered that blood flow dropped and vessel widening decreased. These demonstrate that nitric oxide was actually controlling Sherpa's blood circulation and concluded that Sherpa's high nitric oxide activity increases microcirculatory blood flow, which allows efficient tissue perfusion even under low-oxygen conditions.

The higher level of NO in blood is highly beneficial in hypoxia. Hypoxia triggers hypoxic pulmonary vasoconstriction (HPV)—a reflex where the pulmonary arteries contract to flow blood to better oxygenated areas of the lung. In the case of lowlanders at a high altitude, HPV leads to pulmonary hypertension, which eventually results in right-sided heart failure. However, Sherpas shows a remarkably high cardiac output during exercise without pulmonary hypertension. This is because high levels of nitric oxide can prevent pulmonary hypertension and allow efficient pulmonary circulation.

The high NO levels in blood also affects peripheral microcirculation. Increased NO levels cause systemic vasodilation, which makes blood vessels more permeable and increases blood flow to tissues. Thus, NO levels maximize blood flow to tissue thereby maximize oxygen delivery to tissue even if blood oxygen level is low. Furthermore, sherpa people have greater capillary density in their skeletal muscle which improves oxygen delivery efficiency by reducing diffusion distance from blood vessels to the mitochondria. These vascular adaptations help Sherpa to live at high altitude without harm.

Mitochondrial Adaptation in Sherpa Skeletal Muscle

Genetic adaptation in Sherpas is not restricted to the delivery of oxygen. Sherpa involves the efficiency of mitochondrial energy production under hypoxic conditions. Mitochondria produce adenosine triphosphate (ATP)—which is cellular energy—through oxidative phosphorylation, which requires myriads of oxygen. Under hypoxic conditions, ordinary people in high altitude disrupts mitochondrial function, which causes inefficient ATP production and increased metabolic stress. However, Sherpa's skeletal muscle is able to produce similar amounts of ATP under hypoxic conditions, which implies that their mitochondria works more efficiently. This improved efficiency allows to reduce unnecessary oxygen and keeps ATP levels stable even in hypoxic conditions. As a result, Sherpas can keep their muscles working with lower energetic cost, which is helpful in high-altitude conditions where oxygen availability is rare. Instead of focusing on maximizing power output, their bodies prioritize oxygen efficiency, which allows them to maintain endurance and metabolic stability. One study showed that Sherpas have better efficiency of oxygen utilization and improved muscle metabolism, which means Sherpas' muscles can keep producing energy even at hypoxia (Horscroft et al., 2017).

The PPARA gene is known as the basis of efficient oxygen delivery of Sherpa people. The PPARA gene encodes the peroxisome proliferator-activated receptor alpha (PPAR α), which is a transcription factor for genes involved in β -oxidation of fatty acid in the liver, heart, and skeletal muscle. β -oxidation of fatty acid is the key step for entering fatty acid into the citric acid cycle and producing ATP through fatty acid metabolism. Genomic analysis has shown that Sherpas have a lot of advantageous PPARA alleles, compared to lowlanders. In addition, the advantageous PPARA alleles are also associated with a lower expression of PPARA in Sherpa muscle (Horscroft et al, 2017). This low expression results in decreased fatty acid oxidation, which can be detected by the lower activity of β -oxidation enzyme-3-hydroxyacyl-CoA dehydrogenase and decreased expression of carnitine transporter.

Through reducing dependency on fatty acids and shift toward carbohydrate metabolism from fatty acid metabolism, Sherpa people can use oxygen more efficiently and preserve muscle energy level. This allows them to maintain higher concentration of phosphocreatine and ATP and protects them from the accumulation of harmful lipid intermediates at high altitude (Horscroft et al, 2017). In contrast, in lowlanders at high altitude, incomplete fatty acid metabolism results in accumulation of long-chain acylcarnitines, the indicators of metabolic stress and lipotoxicity.

Sherpa muscle also showed mitochondrial adaptation. In general, it was easily believed that high altitude results in an increase in mitochondrial density to maximize oxygen use. However, Sherpa people possess lower mitochondrial density at muscle than unacclimatized lowlanders. This low density of muscle mitochondria imply that Sherpa mitochondria are more efficient, which might be involved in maximizing oxygen delivery.

Another key mechanism of Sherpa mitochondrial adaptation is mitochondrial coupling, which refers to the great cellular respiration efficiency. During cellular respiration, proton gradients are created in mitochondria through electron transport chains for ATP synthesis. In this process, lots of energy is lost and dissipates as heat due to several reasons. However the main reason is the presence of uncoupling protein (UCP), which leaks oxygen gradients in mitochondria and reduces ATP production efficiency.

However, Sherpa mitochondria produce more mole of ATP per oxygen consumed under hypoxia, compared to lowlanders mitochondria, which is called tighter coupling. This great efficiency is achieved by: reduced uncoupling protein (UCP) activity. Sherpa mitochondria succeeded to optimize complex V stoichiometry and lower ROS-mediated damage in mitochondria inner membrane (Horscroft et al., 2017).

Not only molecular level adaptation, quantity and structural adaptation are also found in Sherpa mitochondria. Sherpa people have greater volume of subsarcolemmal mitochondria—those located directly beneath the cell membrane, in close proximity to capillaries—compared to lowlanders. This unique and efficient arrangement minimizes oxygen dissipation and maximizes oxygen flux inside of blood. These two molecular and structural adaptation in Sherpa muscle suggests presence of co-optimization of oxygen delivery at both the vascular and intracellular levels

How Sherpa adaptation differs from other highlanders

Both Sherpa and Andean highlanders live at high altitude, so they must adapt to survive and work in this environment. However, their main difference is in physiological traits. Sherpa's adaptation differs from other highlanders such as Andean highlanders. Sherpas adapt by using oxygen more efficiently, which means they use the oxygen that already exists. Also, they have slightly high hemoglobin levels compared to others, which allows more fast and smooth flow of their blood. The specific gene EPAS1 gene regulates the body's responses, and prevents massive red blood cell production. This gene plays a crucial role in preventing blood from being thicker, which can result in cardiovascular problems. Sherpa people also have more efficient mitochondria in their muscles, which enables Sherpas to produce energy even in low oxygen conditions like hypoxia. This indicates that Sherpa people do not require extremely high red blood cells to survive in high altitude.

Feature	Sherpa	Tibetan	Andean
Hemoglobin	Slightly elevated	Slightly elevated	Very high
EPAS1 adaptation	Yes (strong)	Yes (strong)	No
NO availability	Very high	High	Moderate
Polycythemia risk	Low	Low	High
CMS prevalence	Very low	Low	High
Metabolic strategy	Carbohydrate-dominant	Similar to Sherpa	RBC-transport dominant
PPARA enrichment	Yes	Partial	No
Altitude (m)	~3,500–5,000+	~3,500–5,000	~3,500–4,000

Table 2. Comparison of physiological and genetic features across major high-altitude human populations. CMS indicates Chronic Mountain Sickness.

One study compares several highlanders such as Sherpa, Tibetans, Andeans, and Ethiopian highlanders (Sharma et al., 2022). Since all of these groups have adapted low-oxygen environments, they have different physiological strategies. Sherpas demonstrated a unique combination of lower concentration of hemoglobin, efficient blood flow, and increased nitric oxide production, while Andean highlanders revealed elevated hemoglobin concentrations to enhance oxygen transport, and Ethiopian showed relatively small amounts of physiological changes (Table 1). This study emphasizes that Sherpas aren't just an example of highlanders and different from other highlanders. Sherpas have a unique combination of physiological changes compared to other highlanders, which makes them one of the most specialized and efficient high-altitude adapted highlanders in the world.

Unlike Sherpa people, Andean highlanders increase oxygen transport in the blood. Andeans have extremely high hemoglobin levels and much more red blood cells. These characteristics allow them to transport more oxygen. However, extremely high hemoglobin levels have the drawback that it can make the blood thicker, which can result in cardiovascular problems and chronic mountain sickness (CMS). Andeans' energy efficiency is not specialized and strong enough like the Sherpas, they rely more on increased oxygen transport and more red blood cell productions (Table 2).

Remarkably, like Sherpa people, Ethiopian highlanders also maintain hemoglobin values but they do not have the EPAS1 variant as shown in Table 2. In spite of the low concentration of hemoglobin concentration, Ethiopian people's oxygen saturation is also high for altitude, which implies the efficient oxygen delivery. Furthermore, unlike Andeans, Ethiopians show minicam polycythemia and do not have chronic mountain sickness. However, unlike Sherpa people, Ethiopians' mitochondrial adaptation is not particularly characterized. One of known characteristics is that their adaptation is based on a genetic variant near the CYB561A3 gene,

which is completely different from the genetic basis of Sherpa. Sherpa and Ethiopian have completely different and independent genetic variant pathways, however, they have similar physiological adaptation outcomes such as low hemoglobin, high oxygen saturation, and absence of chronic mountain sickness. This is a great example of convergent evolution of mitochondria.

Epigenetic modifications developmental plasticity

Although genes such as EPAS1 and EGLN1 are strongly related to high-altitude survival, scientists suggest that the relation between genetic mutations and actual physiological traits like oxygen efficiency and higher levels of nitric oxide still remains unknown. This leads us to think epigenetic modifications must be a main factor in controlling how these genes affect expressing actual physiological traits. One study showed that high-altitude hypoxia can change gene expression through DNA methylation such as oxygen-response pathways (HIF pathways) and metabolism (Zhang et al., 2019). It is also reported that early exposure of hypoxia can lead to long-term physiological traits, including epigenetic changes such as histone modification and DNA methylation. These changes are related to oxygen sensing, metabolism, and vascular development, which means two different individuals with similar genetic backgrounds can have totally different physiological changes. In addition, one research about developmental plasticity suggests that highlanders often have improved oxygen diffusion capacity, and larger lung volumes, which means that the body is not only genetically adapted but also physiologically trained during growth. This kind of unique combination stands out in Sherpa, since they have a combination of inherited genetic traits and environmentally induced gene regulation, which result in a more adaptable and flexible body compared to other highlanders.

Maternal-fetal oxygen transfer and reproductive success

High-altitude environments create challenges for reproduction because low oxygen levels reduce uterine blood flow and oxygen delivery to the placenta, which can lead to tragic results like intrauterine growth restriction, low birth weight, and increased infant mortality. These results also stand out to lowlanders who move to high-altitude environments. However, Sherpa and Andean populations reveal that they are able to have relatively normal birth weights and low infant mortality, which demonstrates that they went through special physiological adaptations and developments (Ahrenes & Singer, 2025). One of the studies suggests that Sherpas have enhanced uterine artery blood flow, which allows more volume of oxygenated blood to reach the placenta and more efficient placental oxygen transfer which means, fetus can receive enough oxygen in hypoxic conditions. Furthermore, traits with improved maternal-fetal oxygen transfer will result in more populations and will be able to maintain stable population growth compared to other groups with low birth succession in high-altitude environments.

Clinical and therapeutic applications

Sherpa's physiological adaptations allow them to function efficiently even in hypoxia conditions without harmful effects that typically appear in lowlanders. Unlike other populations,

Sherpas do not rely on increasing hemoglobin levels, they improve oxygen use without thickening the blood and avoid complications like blood stroke or clots. As previously discussed, Sherpas have special genes such as EPAS1, which controls the hypoxia-inducible factor (HIF) pathway that affects red blood cell production, oxygen metabolism, and blood vessel formation. Sherpa adaptations contain crucial genetic targets for therapy, particularly involving the EPAS1 gene. For this, Sherpa populations have significant clinical implications, since it can be used to improve treatments for conditions like anemia, improve oxygen delivery in ill patients, or help patients adapt to hypoxia. This is already explored in medicine like HIF stabilizers that are used for anemia. Also, Sherpas demonstrate enhanced oxygen use and mitochondrial efficiency. Their efficient energy usage helps to prevent conditions like pulmonary hypertension. This reveals that focusing on improving mitochondria and energy production can help treat diseases like ischemic heart disease. Instead of delivering more oxygen, medicine can help the body use oxygen efficiently like the Sherpa population.

Conclusion

The Sherpa population is one of the most instructive examples of human biological evolution. Over approximately 25,000 to 40,000 years of residence on the Himalayan plateau, which is one of the most extreme environments, natural selection of genetic variants has sculpted Sherpa people's physiological systems as an efficient oxygen delivery machinery. This adaptation solved the problem of chronic hypoxia, permanent in highlands.

With hematological perspective, genetic variation in EPAS1 and EGLN1 changes the HIF signaling pathway to prevent the body from erythropoietic, overproduction of red blood cells. This recalibration of the HIF pathway is crucial, because overproducing red blood cells thickens the blood and results in enormous strain on the hearts. Thus, Sherpa people were able to avoid this erythropoietic problem by sustaining their blood thin and easily flowing. At the level of blood vessels, Sherpa people produce high levels of nitric oxide which contributes to pulmonary and peripheral vasodilation, thereby sustaining a high-flow, and low-resistance blood circulation. This circulatory system is very effective even for people in low-air environments such as highlands.

At the cellular level, Sherpa people have evolved to metabolite glucose instead of fat as their primary fuel through downregulation of PPARA expression. Fatty acid oxidation, which is the core process of fatty acid metabolism, consumes much more oxygen than glucose metabolite, thus the downregulation of PPARA expression is key for oxygen efficient use. Furthermore, mitochondria of Sherpa people are very efficient in maximizing ATP production. The most impressive point of these systematic adaptations from cellular level to tissue level is these adaptations are coupled to each other and make synergetically benefits. For example, the reduced polycythemia decreases viscosity and contributes to efficient blood flow. Then, elevated nitric oxide keeps the vascular conduction which carries blood further and faster. Also, preference of glucose over fatty acid in metabolism minimizes the consumption of oxygen in a cell. Lastly, efficient mitochondria maximize ATP production per glucose. Each adaptation

benefits other adaptations, producing an optimized physiological system that outperforms over the sum of its parts.

Compared to other high-altitude populations, Sherpa possess a very unique evolutionary path. In particular, the Andean population of South America have evolved to produce more red blood cells to compensate for low oxygen conditions, while Sherpa people maintain low red blood cell levels. Overproduction of red blood cells can be thought of as an effective strategy in terms of oxygen delivery but costs a lot in terms of health. Overproduction of red blood cells causes enormous strain on hearts, thereby significantly increasing cardiovascular risk. Sherpa people have evolved in a more safe and efficient way that does not simply rely on pumping more blood or increasing oxygen delivery. A unique genetic path—some of the genetic variants of Sherpa people came from the Denisovans—may contribute to their unique and safe adaptation to hypoxia.

In spite of substantial meta-analysis of this study, there remains research gaps. First, the full polygenic tree of Sherpa adaptation has not been fully identified. Second, with epigenetic perspective, transcriptional and translational regulation remain uncertain. Also, gut microbiome composition, which is a just emerged research area, and maternal-fetal oxygen transfer are still promising area of uncertainty.

In conclusion, studying Sherpa people is more than just understanding one group of people. Studying Sherpa people shows us that human genomes can overcome extreme environmental challenges that took millennia to refine. As DNA sequencing technologies and analytical tools develop, humans just begin to investigate and expand understanding of molecular mechanisms of evolution. In turn, studying Sherpa people helps understanding how natural selection shapes human physiology in the face of extreme environmental demand.

Works Cited

- Ahrens, S., & Singer, D. (2025). Placental adaptation to hypoxia: the case of high-altitude pregnancies. *International Journal of Environmental Research and Public Health*, 22(2), 214.
- Bhandari, S., & Cavalleri, G. L. (2019). Population history and altitude-related adaptation in the Sherpa. *Frontiers in Physiology*, 10, 1116.
- Bhandari, S., Zhang, X., Cui, C., Yangla, Liu, L., Ouzhuluobu, ... & Su, B. (2017). Sherpas share genetic variations with Tibetans for high-altitude adaptation. *Molecular Genetics & Genomic Medicine*, 5(1), 76–84.
- Cumpstey, A. F., Hennis, P. J., Gilbert-Kawai, E. T., Fernandez, B. O., Grant, D., Jenner, W., ... & Feelisch, M. (2020). Effects of dietary nitrate supplementation on microvascular physiology at 4559 m altitude – A randomised controlled trial (Xtreme Alps). *Nitric Oxide*, 94, 27–35.
- Hanaoka, M., Droma, Y., Basnyat, B., Ito, M., Kobayashi, N., Katsuyama, Y., ... & Ota, M. (2012). Genetic variants in EPAS1 contribute to adaptation to high-altitude hypoxia in Sherpas. *PLoS ONE*, 7(12), e50566.
- Horscroft, J. A., Kotwica, A. O., Laner, V., West, J. A., Hennis, P. J., Levett, D. Z., ... & Murray, A. J. (2017). Metabolic basis to Sherpa altitude adaptation. *Proceedings of the National Academy of Sciences*, 114(24), 6382–6387.
- Hu, H., Petousi, N., Glusman, G., Yu, Y., Bohlender, R., Tashi, T., ... & Huff, C. D. (2017). Evolutionary history of Tibetans inferred from whole-genome sequencing. *PLoS Genetics*, 13(4), e1006675.
- Julian, C. G., & Moore, L. G. (2019). Human genetic adaptation to high altitude: evidence from the Andes. *Genes*, 10(2), 150.
- Sharma, V., Varshney, R., & Sethy, N. K. (2022). Human adaptation to high altitude: a review of convergence between genomic and proteomic signatures. *Human Genomics*, 16(1), 21.
- Zhang, X., Yang, Y., & Shi, Q. (2025). DNA methylation in adaptation to high-altitude environments and pathogenesis of related diseases. *Human Genomics*, 19(1), 100.

The Effect of R&D Expenditure on the Labour Force By Atharva Gupta

Abstract

This paper examines the relationship between research and development (R&D) expenditure and labour force size using a global empirical framework. While existing literature extensively documents the role of R&D in driving productivity, economic growth, and innovation, its effects on labour market outcomes remain ambiguous due to the coexistence of job-creating and labour-saving technological forces. Using data for 104 countries from the World Bank's World Development Indicators, the study employs a regression analysis to assess whether R&D expenditure, measured as a percentage of GDP, is associated with variations in total labour force levels. The results reveal a positive and statistically significant relationship between R&D expenditure and labour force size, indicating that countries with higher R&D intensity tend to have larger labour forces. This finding is consistent with innovation-led economic expansion and increased demand for skilled labour, although it does not rule out heterogeneous effects across sectors and skill groups. Overall, the paper contributes to the literature by providing broad cross-country evidence on the labour market implications of R&D investment and highlights the importance of complementary policies in education and skills development.

Keywords R&D expenditure, inflation, labour force

Introduction

Research and development (R&D) expenditure plays a critical role in driving the global economy, influencing both the productive capacity and the innovation of goods and services within an economy. As countries increasingly pursue innovation-led growth strategies, understanding the broader economic implications has become a central policy concern. While a substantial amount of literature documents the positive effects of R&D expenditure on productivity and output, its implications for labour market outcomes remain less clear, particularly with respect to labour force size.

Research and development outcomes are not linear and vary from case to case. On one end of the spectrum, innovation may stimulate economic expansion, firm creation and the development of new industries, thereby increasing labour demand and encouraging labour force participation. On the other hand, technological progress may make firms capital-intensive, substituting capital for labour and reducing employment opportunities in low-skilled or routine occupations. These opposing outcomes suggest that the net impact of R&D on the labour force is an empirical question and may vary across countries, sectors, and stages of development.

Despite extensive research on innovation and employment, relatively few studies examine the direct relationship between R&D expenditure and aggregate labour force size in a broad cross-country setting. Existing empirical research often focuses on specific regions, sectors or labour market indicators such as unemployment or productivity. This paper, however, contributes to the literature by providing a cross-country examination of the association between

R&D expenditure and the labour force size using data from 104 countries from the World Bank. Through the employment of a regression framework, the study aims to establish a correlation that can inform future research and policy-driven discussions on global labour market dynamics.

Understanding how R&D expenditure relates to labour force outcomes, such as productivity and unemployment, is particularly relevant to policymakers seeking to balance technological advancement with labour market development and polarisation. The findings of this paper, therefore, offer insights into the labour market implications of innovation across global economies.

This paper proceeds as follows: Section 2 provides background information on the existing literature, Section 3 highlights the empirical findings, and Section 4 concludes the paper.

Literature Review

Various Methodologies have been used to investigate the relationship between R&D expenditure and labour market outcomes such as regression analysis (Raee et al., 2017; Ciarli et al., 2018; Jabri & ismail, 2010), instrumental variable (Ciarli et al., 2018), descriptive statistics (Wang et al., 2024; Ayaydin & Karaaslan, 2014), panel data regression (Chmielewski et al., 2025; Hecht, 2017), Tobit regression (Tetteh, 2024), Frequency analysis (Malecki et al., 2006), VAR (Buerger et al., 2010), SUR (Kirchhoff et al., 2007), Second generation panel data analysis (Yazgan, 2018), Qualitative data analysis (Emsina, 2014), OLS (Khatun & Afroze, 2016), PMG estimation (Blanco et al., 2016), GMM (Yildirim & Akinci, 2021; Manjon & Merino, 2012), quantile regression (Manjon & Merino, 2012).

The variable used in the literature often varied including net sales (Manjon & Merino, 2012), female labour force (Yildirim & Akinci, 2021), Physical capital stock (Blanco et al., 2016), marginal productivity (Khatun & Afroze, 2016), real GDP growth (Emsina, 2014), human capital (Jajri & ismail, 2010), sustainable economic growth (Yazgan, 2018), Liquidity (Ayaydin & Karaaslan, 2014), Purchasing power parity (Kirchhoff et al., 2007), Patents (Buerger et al., 2010), Capital share (Hecht, 2017), Quality of education (Malecki et al., 2006), Laour productivity (Tetteh, 2024), Unemployment (Chmielewski et al., 2025), AI development (Wang et al., 2024), Deregulation index (Giustini, 2021), Degree of labour market routinization (Cirali et al., 2018), GDP/Labour force (Raee et al., 2017), R&D expenditure (Raee et al., 2017).

Many studies have focussed on various countries and regions such as the UK (Cirali et al., 2018), China (Wang et al., 2024), The EU (Chmielewski et al., 2025; Emsina, 2014; Majon & Merino, 2012), Ghana (Tetteh, 2024), USA (Malecki et al., 2006; Kirchhoff et al., 2007; Blanco et al., 2016), Germany (Buerger et al., 2010), Turkey (Ayaydin & Karaaslan, 2014), Malaysia (Jajri & Ismail, 2010), Asia (Khatun & Afroze, 2016), Middle income countries (Yildirim & Akinci, 2021).

R&D expenditure is widely recognised as a primary driver of productivity and national competitiveness. For example, Raee et al. (2017) found that government investment in R&D is linked to higher labour productivity growth; similarly, Tetteh (2024) observed that in Ghana's manufacturing sector, R&D significantly improved innovation and productivity. This positive

relationship extends to broader economic implications; for instance, Yazgan (2018) demonstrated that in OECD countries, R&D investment has boosted economic growth.

In addition to aggregate growth effects, the literature highlights important differences in how R&D influences labour markets and income distribution. Moreover, Hecht (2017) shows that while R&D raises output and productivity at a global scale, technological change is largely labour saving, leading to uneven employment effects across sectors. This is supported by Ciarli et al. (2018), who found that R&D intensifies labour market polarisation in the United Kingdom (UK) by expanding high skilled employment in routineised areas. In contrast, Wang et al. (2024) find that AI driven R&D in China upgraded the employment structure by reducing low skilled jobs and expanding medium and high skilled employment, suggesting that labour market outcomes depend heavily on industrial structure and skill composition.

Several studies also distinguish between types and locations of R&D investment. Chmielewski et al. (2025) find that in EU countries, business R&D delivers faster economic benefits than government R&D, which may have delayed effects and even widen gender inequality if poorly targeted. University-based R&D, however, is shown to affect growth indirectly. Kirchhoff et al. (2007) and Blanco et al. (2016) find that university R&D stimulates firm creation and regional spillovers, thereby generating employment and productivity gains through entrepreneurship rather than direct job creation. Spatially, Malecki et al. (2006) and Buerger et al. (2010) show that R&D remains concentrated in large urban centres, reinforcing regional disparities despite some evidence of spillovers.

Overall, the literature suggests that while R&D is a robust engine of productivity and growth, its employment and distributional effects are heterogeneous and highly dependent on complementary policies in education, labour market regulation, and regional development.

Table 1. An Extensive Review of Literature on the Effect of R&D Expenditure on the Labour Force

Authors	Year	Key Variable	Region/ country studied	Methodology	Findings	Data Period
Raee et al.	2017	GDP/labour force, PPP	Global	Regression	Government investment in tertiary education and R&D increases labour productivity growth.	2006–2015
Ciarli et al.	2018	Employment, labour routinisation	UK	Regression, IV	R&D intensifies job polarisation, increasing demand for reskilling and retraining.	2001–2012
Giustini	2021	Labour flexibility, digitalisation	Global	Meta-analysis	Digitalisation makes R&D work more flexible but also more precarious.	NA
Wang et al.	2024	AI development, employment structure	China	Descriptive statistics	AI reduces low-skilled jobs while expanding medium- and high-skilled employment.	2010–2019
Chmielewski et al.	2025	Unemployment, gender pay gap, FDI	EU	Panel data regression	Business R&D boosts growth quickly, while public R&D has delayed and unequal effects.	2013–2022
Tetteh	2024	Innovation, labour productivity	Ghana	Tobit regression	R&D significantly improves productivity and innovation in manufacturing.	2013
Malecki et al.	2006	Job alternatives, education quality	USA	Survey frequency analysis	Skilled labour and universities attract R&D, which remains spatially concentrated.	1992
Hecht	2017	R&D, capital share	Global	Panel data analysis	R&D raises output and labour productivity through labour-saving technical change.	1992–2013
Buerger et al.	2010	Patents, employment, growth	Germany	VAR	Patent growth predicts employment and R&D growth in select industries.	1999–2005
Kirchhoff et al.	2007	Innovation, firm creation	USA	OLS, SUR, 2SLS	University R&D stimulates firm formation and employment via spillovers.	1990–1999
Ayaydin & Karaaslan	2014	R&D intensity, firm performance	Turkey	Descriptive stats, dynamic panel	Higher R&D intensity improves firm profitability, especially in high-tech sectors.	2008–2013
Yazgan	2018	R&D investment, growth	OECD	Panel data analysis	R&D drives growth more strongly in high-income OECD countries.	1996–2015
Jajri & Ismail	2010	GDP, human capital	Malaysia	Cobb–Douglas, regression	Physical capital dominates growth, but skilled labour boosts productivity.	1981–2007

Emsina	2014	Productivity, competitiveness	EU	Mixed quantitative–qualitative	Productivity growth supported post-crisis recovery and competitiveness.	2006–2013
Khatun & Afroze	2016	GDP, labour productivity	Asia	Cobb–Douglas, OLS	Labour contributes more than capital to GDP, especially with education investment.	1990–2014
Blanco et al.	2016	R&D, GDP growth	USA	DOLS, PMG	R&D increases long-term productivity with strong inter-state spillovers.	1963–2007
Yildirim & Akinci	2021	Education, female labour force	Middle-income countries	GMM	Education raises female labour participation and supports growth.	2001–2016
Manjón Merino	2012	R&D intensity, firm growth	Europe	OLS, GMM, quantile regression	R&D is critical for firm growth and survival, especially in high-tech sectors.	2003–2007

Empirical Analysis

Data

This empirical study utilises cross-country data from the World Bank's World Development Indicators (WDI) to examine the relationship between R&D expenditure and the labour force. The dataset is based on the most up-to-date and extensive data available, with observations extending up to 2022, which represents the latest year with broad global coverage of R&D expenditure and labour force indicators.

The dataset includes a wide range of individual countries as well as World Bank aggregate classifications (such as regions, income groups, and demographic groupings). To ensure consistency, the empirical analysis retains individual countries only, while aggregate entities (e.g. *World*, *OECD members*, *High income*, and regional aggregates) are excluded from the regression sample. After this filtering process, the final sample consists of 104 countries shown in Appendix 2.

R&D expenditure is measured as gross domestic expenditure on research and development as a percentage of GDP, encompassing spending by the business, government, higher education, and non-profit sectors. The labour force variable captures the total economically active population aged 15 and above, including both employed individuals and those actively seeking work. For ease of interpretation, the labour force variable is scaled by 10^6 .

The dataset is unbalanced, reflecting differences in reporting across countries and years. The country sample spans all major World Bank regions, allowing the analysis to capture heterogeneous innovation and labour market relationships across different levels of economic development. A complete list of countries included in the study is provided below.

Descriptive Statistics

Table 2. Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Labour Force (x10 ⁶)	235	160.376	471.032	0.031	3560.824
R&D Expenditure	109	1.374	1.159	0.041	5.783

Table 2 provides descriptive statistics for the labour force and R&D expenditure. The labour force value is divided by 10^6 . The R&D expenditure has 109 observations with a mean of 1.374 and a standard deviation of 1.159, ranging from 0.041 to 5.783. The labour force has a mean of 160.376 and a standard deviation of 471.032, ranging from 0.031 to 3560.82

Model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Where Y_i is the dependent variable for observation i , which refers to the labour force. β_0 is the constant term, representing the expected value of the variable when the independent variable is 0. β_1 is the coefficient for the independent variable, which is R&D expenditure. The coefficient shows how much the dependent variable changes when the independent variable changes by one unit. ε_i is the error term, which represents the difference between the actual value and the predicted value from the model. The study utilises the linear regression model based on data from the World Bank.

A Regression analysis is a statistical method used to examine the relationship between a dependent variable and independent variable(s). This type of analysis helps us understand how changes in the independent variables correlate with changes in the dependent variable. Moreover, a regression analysis can be used for prediction, forecasting and identifying trends and patterns by fitting a mathematical model to a data set (Gujarati, 2021) .

Based on the findings from Table 1, the estimated regression equation can be expressed as:

$$\text{Labour Force}_i = 1.856 + 7.404 * \text{R\&D Expenditure}_i + \varepsilon_i$$

1.856 represents the predicted value of the labour force when R&D Expenditure is 0.

7.404 indicates the magnitude of the change in labour force for every unit increase in R&D Expenditure.

Findings

Figure 1. Scatterplot Graphics between Labour Force and R&D Expenditure

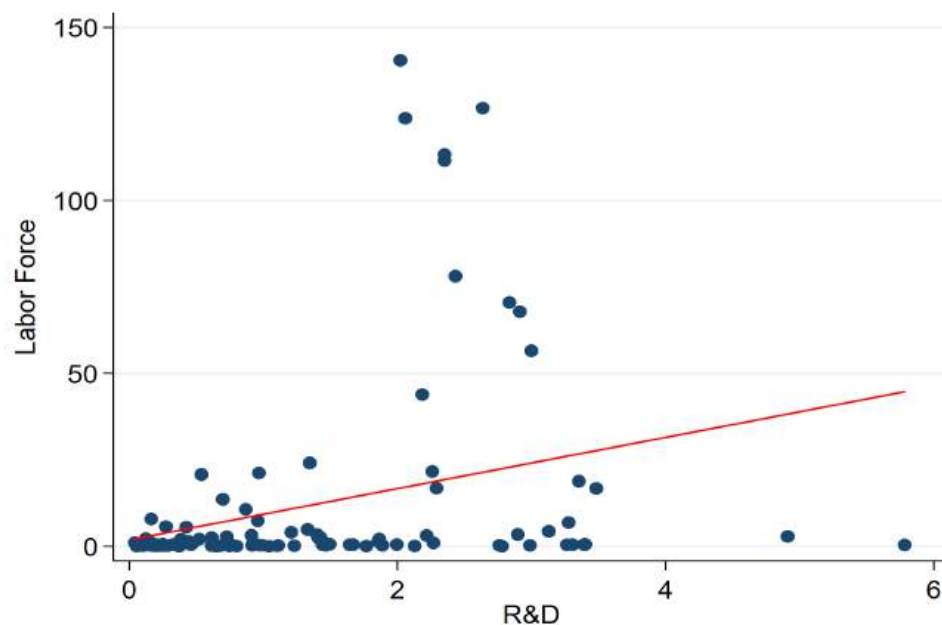


Figure 1 illustrates a scatterplot of the relationship between Labour Force and R&D Expenditure. R&D Expenditure positively correlates with the labour force, suggesting that a higher level of

R&D Expenditure contributes to a higher level of Total Labour Force. This positive correlation could have occurred due to increased job creation as a result of higher R&D expenditure and innovation within the economy. Moreover, the quantity demanded for skilled labour increases as new job opportunities that require skilled labour are created, fostering overall economic growth.

This positive relationship can also be explained through the lens of Foreign Direct Investment (FDI). As markets increase their R&D expenditure, they become more attractive to foreign direct investment, which brings in new job opportunities and new capital that can increase the productivity of an economy. Overall, an increase in R&D expenditure boosts economic growth through reduced unemployment and an increase in the output of total goods and services within an economy, leading to an increase in aggregate demand.

Table 3. Regression Analysis

VARIABLES	(Model 1)
	Labour Force
R&D expenditure	7.404*** (2.670)
Constant	1.856 (2.067)
Observations	104
R-squared	0.090

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3 presents the results of the regression models where the dependent variable is the labour force and the independent variable is R&D expenditure. The paper uses key values for defining the significance level of the estimated parameter. We indicate the statistical significance of the result at the 1% level (***), the 5% level (**) and the 10% level (*). The sign in front of the beta coefficient suggests whether an explanatory variable has a statistically positive or negative impact on the dependent variable. A minus (-) represents a negative impact, whilst a plus (+) represents a positive impact. In Table 2, R&D expenditure is positively associated with the labour force. The coefficient for R&D expenditure is 7.404, which is statistically significant at the 1% level (p-value<0.01). This coefficient suggests that for every 1% increase in R&D expenditure, the total labour force is expected to increase by approximately 7.404 units (Where units are scaled by 10⁶).

Conclusion

This study examines the relationship between research and development (R&D) expenditure and the size of the labour force using the data from 104 countries. The empirical analysis suggests a positive relationship between these 2 variables, indicating that a higher level

of research and development expenditure is linked to a higher level of labour force. Specifically, the regression result indicated that for every 1% increase in R&D expenditure, the total labour force is expected to increase by approximately 7.4×10^6 units. The result highlights that research and development expenditure is a key indicator for expanding the labour market.

As R&D expenditure continues to rise, so does the demand for skilled labour. Therefore, policymakers and governments should invest in education and nuanced skills development to complement innovation policies and improve the quality of human capital over the long run. Moreover, due to an everlasting increase in R&D expenditure, skill-biased technological change is inevitable, causing higher productivity sectors to emerge and drive global output. Without substantial investment in education, labour supply will not be equal to labour demanded and therefore structural unemployment may arise. Policymakers should also differentiate R&D support strategies differently, taking into account the specifics that come with each sector.

Future studies should take a more extensive time period into account. By doing so, a relationship between R&D expenditure and labour markets may have more conviction regarding whether this is a correlation or causation. Additionally, a more nuanced and advanced methodology could be used, such as a panel data analysis, in order to control both the cross-sectional and the time dimension effects. Future studies could also take the time period of COVID-19 into account and see whether the systematic risk period had any significant effect on this relationship.

Works Cited

- Al Raei, M., Ritzen, J., & de Crombrugghe, D. (2017). Innovation policy & labour productivity growth: Education, research & development, government effectiveness and business policy. *Innovation*, 2017, 019.
- Ciarli, T., Marzucchi, A., Salgado, E., & Savona, M. (2018). The effect of R&D growth on employment and self-employment in local labour markets.
- Giustini, D. (2021). The impact of labour market trends on the employment of R&D personnel: a literature review. Higher School of Economics Research Paper No. WP BRP, 117.
- Wang, X., Chen, M., & Chen, N. (2024). How artificial intelligence affects the labour force employment structure from the perspective of industrial structure optimisation. *Heliyon*, 10(5).
- Chmielewski, W., Postuła, M., & Gawkowski, K. (2025). Linking R&D Expenditure to Labour Market and Economic Performance: Empirical Evidence from the European Union. *Sustainability*, 17(14), 6595.
- Tetteh, C. K. (2024). R&D, innovation and labour productivity: Evidence from the manufacturing sector in Ghana. *Scientific African*, 24, e02136.
- Malecki, E. J., & Bradbury, S. L. (1992). R&D facilities and professional labour: labour force dynamics in high technology. *Regional Studies*, 26(2), 123-136.
- Hecht, J. (2018). Research and development and labour productivity: do high-tech firms exhibit labour-or capital-saving technical change?. *Applied Economics*, 50(16), 1790-1811.
- Buerger, M., Broekel, T., & Coad, A. (2012). Regional dynamics of innovation: Investigating the co-evolution of patents, research and development (R&D), and employment. *Regional Studies*, 46(5), 565-582.
- Kirchhoff, B. A., Newbert, S. L., Hasan, I., & Armington, C. (2007). The influence of university R & D expenditures on new business formations and employment growth. *Entrepreneurship theory and practice*, 31(4), 543-559.
- Ayaydın, H., & Karaaslan, I. (2014). The Effect of Research And Development Investment On Firms' financial Performance: Evidence From Manufacturing Firms In Turkey. *Bilgi ekonomisi ve yönetimi dergisi*, 9(1), 23-39.
- Yazgan, Ş., & Yalçinkaya, Ö. (2018). The effects of research and development (R&D) investments on sustainable economic growth: Evidence from OECD countries (1996-2015). *Review of Economic Perspectives*, 18(1), 3-23.
- Jajri, I., & Ismail, R. (2010). Impact of labour quality on labour productivity and economic growth. *African Journal of Business Management*, 4(4), 486.
- Auzina-Emsina, A. (2014). Labour productivity, economic growth and global competitiveness in post-crisis period. *Procedia-Social and Behavioral Sciences*, 156, 317-321.
- Blanco, L. R., Gu, J., & Prieger, J. E. (2016). The impact of research and development on economic growth and productivity in the US States. *Southern Economic Journal*, 82(3), 914-934.

Yıldırım, D. Ç., & Akinci, H. (2021). The dynamic relationships between the female labour force and the economic growth. *Journal of Economic Studies*, 48(8), 1512-1527.

García-Manjón, J. V., & Romero-Merino, M. E. (2012). Research, development, and firm growth. Empirical evidence from European top R&D spending firms. *Research Policy*, 41(6), 1084-1092.

Gujarati, D. N. (2021). *Essentials of econometrics*. Sage Publications.

Appendix

Appendix 1. Source of the Data

Indicator Name	Variable	Definition	Source
R&D expenditure	Research and development expenditure (% of GDP)	Gross domestic expenditures on research and development (R&D), expressed as a percentage of GDP. They include both capital and current expenditures in the four main sectors: Business enterprise, Government, Higher education and Private non-profit. R&D covers basic research, applied research, and experimental development.	WDI World Bank
Labour Force	Labour force, total	The labour force comprises people aged 15 and older who supply labour for the production of goods and services during a specified period. It includes people who are currently employed and people who are unemployed but seeking work, as well as first-time job-seekers.	WDI World Bank

Appendix 2. List of countries present within the data set

North America	South America	Europe	Asia	Oceania
Canada, USA	Argentina, Chile, Paraguay, Peru, Uruguay	Armenia, Austria, Azerbaijan, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Moldova, Netherlands, North-Macedonia, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovak-Republic, Slovenia, Spain, Sweden, Switzerland, Ukraine, UK	China, Hong Kong SAR, China, Macao SAR, China, Japan, Korea, Kazakhstan, Kyrgyz Republic, Mongolia, Myanmar, Pakistan, Thailand, Turkmenistan, Uzbekistan, Viet Nam, Iran, Iraq, Israel, Oman, Qatar, Saudi Arabia, UAE	Australia, New Zealand

How Airport Design and National Airlines Shape Soft Power and National Identity

By Austin Cao

Abstract

In 2024, airlines carried an estimated five billion passengers, suggesting air travel plays an important role in transporting passengers from one country (and national identity) to another. This paper investigates the question, “In what ways do national airlines and airport design communicate a country’s image, national identity and soft power?” To answer that question, I conduct an interdisciplinary literature review, synthesizing economic, sociological, and historical views on national identity and air travel. I conclude that national identity and soft power are both key sources for an airline to become “flying monuments” for their nations, and an important part of nation branding.

Introduction

In 2024, roughly five billion passengers were transported by airlines worldwide, an increase from 2023, during which 4.4 billion passengers were carried (Air Transport Action Group). This statistic shows the importance of air travel among civilians today, as it is considered the safest transportation method in the world. In comparison, in 1975, around 432 million passengers were carried by all commercial airlines worldwide (World Bank). Air travel is also important for countries: this paper aims to answer the research question, “In what ways do national airlines and airport design communicate a country’s image, national identity, and soft power?” Answering this question reveals how a national airline represents a country’s national pride, and sheds light on two important terms, national identity and soft power, which represent how national airlines and airport designs symbolize a country.

National identity refers to a person’s sense of belonging to a specific nation, culture, history, and values that bind the members of a country together (Huddy and Khatib; Levendusky). While soft power refers to the ability to influence others through attention (Nye), in aviation, soft power uses air travel to boost airlines’ image. Having knowledge of national pride, national identity, and soft power is important as these terms are essential for understanding foreign policy and influences of countries globally. These terms are closely connected, with national identity and pride being a foundation of soft power (Nye).

In this paper, I conduct a literature review focusing on important concepts such as national identity and soft power. Both terms play an important role in shaping a national airline’s pride in their country. Airlines not only fly passengers from one place to another but also help connect geographies and the people that belong to these geographies who may be disadvantaged by linking them to the more developed world.

Based on research, I have found that different airlines represent national identity and soft power in their own ways. Not only do they serve as a symbol for airlines, they were important symbols during wartime. Additionally, national identity and soft power are reflected in airport

designs. This shows that not only airports serve as an important transportation hub, they also serve as a place where countries show their identity and how they are viewed by passengers.

How Airlines Represent a Country's Identity and Global Image

In the aviation world, national identity plays a significant role. It represents a symbol of national pride, especially for national airlines in the Middle East such as Emirates, Qatar, and Etihad, which serve as the gateway to the world with destinations to all continents (Legnani). National airlines demonstrate soft power by connecting their country's brand and culture to the rest of the world. An important part of the soft power of airlines is an airline's livery, since it typically features national symbols, flags, and colors, which causes a country's culture to be known to the world. The tailfin is one of the most essential parts of an airplane as it serves as a symbol for the airline in addition to keeping the plane stable in the air. Flag carriers can benefit from their status of being the national airline, as successful national airlines typically have extensive global reach and a geographical advantage. Airlines like Emirates and Qatar are headquartered in the Middle East, making these airlines appealing for transit passengers. Additionally, Middle East airports have among the most destinations worldwide.

Turkish airlines (TK), the national carrier of Turkey, is an example of how a country uses a national airline to shape its identity. The airline is a major gateway to the world, with over 290 destinations worldwide (Anaz). Additionally, TK has become the face of Turkey's transportation (Anaz). The design of TK's planes is a wild goose for its livery, symbolizing the airline's commitment to high altitude and long distance flights. TK uses its good customer service to influence customers' perceptions of the country itself by treating their passengers with warmth and hospitality, which is inspired by the characteristics of Turkish people. Additionally, TK is known for its gourmet dining experience and onboard "flying chefs," which showcases their quality Turkish cuisine. This allowed TK to become a major facilitator of Turkey's international reputation.



Figure 1: An turkish airlines B777-300 (google image)

Qatar Airways is another good example, as they influence customers' opinions of Qatar by providing passengers with a high level of consistency and service across all flights. They recognize that all passengers are different and therefore allow passengers to decide how they wish to communicate to the crews in order to make their in-flight experience comfortable. Passengers can decide how often they communicate with crews, and if they feel reserved, some airlines have an option on their entertainment screens to request assistance, instead of calling for service manually (Future Travel Experience). Additionally, the design of Qatar Airways' livery is an Arabian Oryx, which is the national animal of Qatar. It symbolizes the airline's, and, by extension, the country's, strength and spirit. Additionally, the logo of Royal Jordanian Airlines, which is a gold crown, symbolizes the airline's connection to the Hashemite Kingdom of Jordan, with the gold color representing both luxury and modernity.

Significance Of National Identity During Wartimes

Turkish Airlines and Qatar Airways serve as examples of how nations now use aviation to promote their national identities. This tendency is not new, though, and it is not limited to commercial aviation. In the past, countries have employed military aircraft to promote their national identities. Nose Art, which is the design on a fuselage of an aircraft, was an important war symbol during WWII (Curatola). This concept started during WWII, and countries such as the USA, Germany, Britain, France, Australia, and Canada, have been involved in nose art. The designs of nose art featured cartoon characters, symbols such as animals, and many more. The purpose of it is to provide relief to crews during wartimes through self-expression in the art (Barber).

Soft Power In Airport Design

Not only do airlines have soft power, but airport design does too. The more iconic, futuristic, and modern an airport design is, the more positive impressions travelers will have of the airport, as modern airports typically include efficient security checkpoints, automated passport control, etc. For example, modern airports such as Doha, Singapore, and Hong Kong have consistently been ranked among the top airports in the world (Skytrax), reflecting the high praise that these major airports get from passengers. The most notable airports with an futuristic design include Beijing Daxing Airport's starfish design, which aims to increase efficiency by minimizing walking distances to gates. Singapore Changi is another example, as it features the world's largest indoor waterfall, which is located inside the Jewel, a shopping mall inside the airport (Volume Zero). With this waterfall, the airport creates a calming experience for passengers in a high traffic environment. Overall, these modern airports show that futuristic airport design isn't about looking impressive but rather about passengers' experience of the airport based on comfort and efficiency.



Figure 2: Indoor waterfall in the Singapore airport(Raut)

Conclusion

In conclusion, national identity and soft power play an important role in national airlines' image of a country. Airlines and airports act as important symbols of a nation, shaping their image and influence to the world through their cultural representation and service to the greater civilians. An idea for future research on aviation would be to explore the role of aircraft interiors in national identity and soft power, which could analyze how cabin design, service style, and passenger experience contributes to an airline's national image beyond the exterior design of an aircraft. Lastly, in a future research project, researchers could collect data by interviewing airline staff, airport designers, and passengers to get direct insights from these professionals.

Works Cited

- “Air transport, passengers carried.” World Bank,
<https://data.worldbank.org/indicator/IS.AIR.PSGR>. Accessed 9 February 2026.
- Alves, Glynda. “Twist in the Tail: The Significance of Airline Liveries and Tailfins.” *The Economic Times*,
economictimes.indiatimes.com/industry/transportation/airlines/-aviation/twist-in-the-tail-the-significance-of-airline-liveries-and-tailfins/articleshow/105217186.cms?from=mdr. Accessed 9 Feb. 2026.
- “AP Human Geography Review: National Identity.” Fiveable,
fiveable.me/key-terms/ap-hug/national-identity. Accessed 9 Feb. 2026.
- Barber, Martin. “How Nose Art and Pin-Ups Boosted WWII USAAF Troops’ Morale.” *BBC*,
<https://www.bbc.com/news/uk-england-19288744>. Accessed 9 February 2026.
- Curatola, John. “Decoding WWII Plane Nose Art.” *The National WWII Museum*,
<https://www.nationalww2museum.org/war/articles/decoding-wwii-plane-nose-art>. Accessed 9 February 2026.
- Diskin, Eben. “Turkish Airlines’ Flying Chefs Are Taking Inflight Fine Dining to the Next Level.” *Matador Network*,
<https://matadornetwork.com/read/turkish-airlines-flying-chefs/>. Accessed 9 February 2026.
- “Facts & Figures.” Air Transport Action Group,
atag.org/facts-figures#:~:text=Connectivity&text=In%202023%2C%204.4%20billion%20passengers,2024%20is%205%20billion%20passengers.&text=In%202023%2C%2035.3%20million%20scheduled%20commercial%20flights%20took%20place%20worldwide. Accessed 9 Feb. 2026.
- Ghuge, Samiksha, “Flying into the Future: Most Futuristic Airports Designed Around the World.” *Volume Zero*,
<https://volzero.com/news/view/flying-into-the-future-most-futuristic-airports-designed-around-the-world>. Accessed 9 February 2026.
- “How Qatar Airways is Empowering Passengers to Take Control of Their Journey Through Digital Means.” *Future Travel Experience*,
<https://www.futuretravelexperience.com/2021/05/how-qatar-airways-is-empowering-passengers-to-take-control-of-their-journey-through-digital-means/#:~:text=We%20are%20well%20known%20and,and%20expectations%20of%20the%20event>. Accessed 9 February 2026.
- Legani, Matteo. “‘Global’ Companies for Real: The Magnificent Nine that Fly to all Six Continents.” *The Flight Club*,
<https://www.theflightclub.it/en/2024/01/companies-fly-all-over-the-world-six-continents-such-as/>. Accessed 9 February 2026.
- McClory, Jonathan. “Love From Above: the Soft Power of Aviation.” *Aviation Benefits*,

- https://aviationbenefits.org/media/13961/ATAG_AviationBenefits2014_ESSAY_Love-from-above.pdf#:~:text=On%20a%20global%20level%2C%20aviation%20facilitates%20international,hard%2C%20power%20drives%20change%20for%20the%20better. Accessed 9 February 2026.
- Nye, Joseph S. "Soft Power and Morals in U.S. Foreign Policy." Sasakawa USA, 8 June 2021, spfsa.org/publications/soft-power-and-morals-in-u-s-foreign-policy/#:~:text=Soft%20power%20is%20the%20ability,can%20attract%20others%20by%20example.
- "Qatar Airways Logo." LOGOS-WORLD, <https://logos-world.net/qatar-airways-logo/>. Accessed 9 February 2026.
- "Royal Jordanian Logo." LOGOS-WORLD, <https://logos-world.net/royal-jordanian-logo/>. Accessed 9 February 2026.
- Raut, Avni. "Take a Look at the World's Tallest Indoor Waterfall at the Jewel Changi Airport, Singapore." AD, <https://www.architecturaldigest.in/content/singapore-changi-airport-worlds-tallest-indoor-waterfall-safdie-architects/>. Accessed 9 February 2026.
- Smith, Tom W., and Lars Jarkko. "National Pride: A Cross-National Analysis." CNR19 National Pride - A Cross-National Analysis.Pdf, May 1998, [gss.norc.org/content/dam/gss/get-documentation/pdf/reports/cross-national-reports/CNR19%20National%20Pride%20-%20A%20cross-national%20analysis.pdf](https://gss.norc.umd.edu/content/dam/gss/get-documentation/pdf/reports/cross-national-reports/CNR19%20National%20Pride%20-%20A%20cross-national%20analysis.pdf).
- "World's Top 100 Airports 2025." Skytrax World Airport Awards, <https://www.worldairportawards.com/worlds-top-100-airports-2025/>. Accessed 9 February 2026.

T-Visa Backlogs Put Trafficking Survivors at Risk By Sabina Kovach

Abstract

Victims of human trafficking face an uphill battle when they escape their abusers, particularly if they are not located in their home country. In response, many countries have created legislation designed to provide visas or residence permits to trafficking survivors to provide victims a recovery period after their escape. The United States created the T-visa system to provide trafficking survivors temporary residency, work authorization, and a path to permanent residency. Despite the positive intentions of the T-visa program, backlogs in application processing have created long waits for T-visas, leaving foreign survivors without legal protections during a particularly vulnerable time. In this article, I outline the primary issues with T-visas in the U.S. today and policies that could potentially fix program delays.

Context

T-visas were created in 2000 by the United States Congress to serve as a form of legal protection for victims of human trafficking. A T-visa is a form of temporary protection that can be given to survivors of “severe form[s] of human trafficking that occurred in or on the way to the U.S.” (U.S. Citizenship and Immigration Services). They were created through the Victims of Trafficking and Violence Protection Act and enable survivors of trafficking to escape dangerous situations without facing the risk of deportation and provide them a legal path to U.S. citizenship. By acquiring a T-visa, a person gains the ability to legally live and work within the U.S. for up to 4 years in addition to protection from deportation.

Recently, there has been a sharp increase in T-visa applications with “applications filed by victims of trafficking [rising] by an astounding 801%” between the years 2021 and 2024, leading to a severe backlog within the system (Gonzalez). However, due to this spike in application, the T-visa application process has become prolonged, leading to a rise in rejections and increased wait times. Because of this, thousands of T-visa applicants are left exposed and vulnerable and are forced to wait up to several years while their cases are pending. Many of these survivors do not receive work permits or any protection from possible deportation during this waiting period. This issue will continue to worsen if action is not taken soon.

Stakeholders & Impact

T-visa backlogs are a significant issue because they leave previously exploited people vulnerable to experience further trauma and hardships. Wait times are “taking now over 20 years” for their application to be processed (Do). While there has been action taken in recent years, “they don’t fully shield people from the risk of deportation while they wait” (Singh). Due to the applicants’ situations, these wait times can be dangerous to their mental or physical health and can impact their ability to care for their families. One applicant said that she is “living in the in-between” as she awaits her visa decision (Barclay). These wait times create a situation where

applicants may be deported and are unable to work—a dangerous combination for someone already living in a precarious situation.

What's Been Done

In response to the rise in applications, the “USCIS introduced a ‘bona fide determination’ in August 2024, a checkpoint meant to identify strong cases early” (Singh). This alerts the people handling T-visa applications of more important cases, allowing survivors with strong cases a shorter waiting period for their case to be processed. This policy also grants temporary benefits to survivors, such as work authorization, and reduces the risk of deportation. In turn, this helps reduce the risk survivors face during that wait time and improves victims’ ability to gain economic independence. However, this policy does not fully solve the application backlog issue, as not all applicants are able to receive this rapid processing and determination. As a result, many victims deal with considerable uncertainty.

Why It’s Not Enough

Even with the bona fide determination put in place, current T-visa backlogs and victim wait times are unacceptable. Due to an increase in the amount of staff being fired from the USCIS under the Trump administration, the backlog has gotten worse rather than better.

To make matters worse, the number of T-visa applicants has surged, and the system has become less efficient. In fact, “between October 2023 and September 2024, the agency received a record 15,332 T-visa applications and approved 3,786. About 600 applications were denied, and the backlog grew from about 9,000 applications the year prior to more than 20,000” (Barclay). Rising denials and a backlog of applicants means that the process has never been worse for applicants. These increasing wait times “have taken a toll” on those who have submitted applications (Barclay).

Policy Proposal

A major contributor to the increasing backlog within the T-visa system has been the lack of staff available to review applications. Action needs to be taken to hire more people that are well equipped to handle the often complex cases involving numerous examples of trafficking and trauma. This is especially critical when the number of applications has been exceeding the cap of 5,000 visas for years now. Many groups advocating for the protection of trafficking survivors have already stated that this cap should be increased or removed altogether in order to ensure that those eligible are not put onto a waitlist after their cases have been approved.

In addition to hiring extra personnel, further protection for the survivors needs to be provided. Bona fide determination is a solution already being implemented to give survivors some protection while they wait on the actual T-visa by providing benefits such as authorization to work. The policy has already shown great promise; however, it is not available to all applicants. If another policy similar to this were implemented for all applicants, this would

greatly increase the safety of thousands of survivors' lives by allowing all victims to achieve economic independence while waiting for T-visa hearings.

Conclusion

The T-visa system was created to protect survivors, but with a heavy backlog in recent years, many applicants are left with little safety or stability. If meaningful action is not taken, the cycle of exploitation and instability will continue while victims wait many years for meaningful legal protections. Significant change needs to occur within this system to match the rapidly increasing number of T-visa applicants.

Works Cited

- Barclay, Mel Leonor. "The Protective Visas That May Never Come." The 19th, 3 Sept. 2025, <https://19thnews.org/2025/09/u-t-visas-victims-violence-immigrants-women/>.
- Do, Phi. "More Human-Trafficking Survivors Are Seeking Visas but Face Longer Waits and Risk Deportation." Los Angeles Times, 23 Aug. 2025, <https://www.latimes.com/world-nation/story/2025-08-23/more-human-trafficking-survivors-are-seeking-visas-but-face-longer-waits-and-risk-deportation>. World & Nation.
- Gonzalez, Erika. "T Visa Applications: Trends and Projections." CAST Training and Technical Assistance Program, 6 Jan. 2025, http://casttta.nationbuilder.com/the_rising_trend_of_t Visa_applications_fy2024.
- Singh, Shashank. "More Trafficking Survivors Apply for T Visas amid Longer Waits and Deportation Risks - VisaVerge." Visa Verge, 23 Aug. 2025, <https://www.visaverge.com/news/more-trafficking-survivors-apply-for-t-visas-amid-longer-waits-and-deportation-risks/>.
- U.S. Citizenship and Immigration Services. "USCIS Restores Integrity to the VAWA, T Nonimmigrant, and U Nonimmigrant Programs After Suspected Fraud." U.S. Citizenship and Immigration Services, 22 Dec. 2025, <https://www.uscis.gov/newsroom/alerts/uscis-restores-integrity-to-the-awa-t-nonimmigrant-and-u-nonimmigrant-programs-after-suspected>.

A Prototype Python Workflow for Automated Extraction and Visualization of Quantum Chemical Data from Gaussian Outputs By Abhiram Pannala

Abstract

Quantum chemical calculations provide detailed insights into the structural and electronic properties of molecular systems. However, computational packages often generate large and complex output files that require manual examination to extract relevant information. Software such as Gaussian 09 produces extensive log files containing optimized geometries, electronic energies, vibrational frequencies, and molecular orbital energies. Extracting these parameters manually can be time-consuming, particularly when multiple molecules are analyzed. This study presents a prototype workflow integrating computational chemistry calculations with automated data extraction using the programming language Python. Density Functional Theory (DFT) calculations were performed on a set of small molecular systems to obtain optimized structures, vibrational frequencies, and frontier molecular orbital energies. A Python-based parser was then developed to extract selected descriptors from Gaussian output files and organize them into structured datasets. The extracted data were used to generate visual representations such as infrared spectra and HOMO-LUMO energy diagrams. The results demonstrate how simple automation tools can significantly streamline the analysis of quantum chemical outputs. Although the present workflow represents a prototype implementation applied to a small molecular dataset, it illustrates how programming techniques can assist in managing computational chemistry data. The approach highlights the value of combining computer science and theoretical chemistry to improve efficiency and accessibility in molecular modeling research and education.

Keywords Density Functional Theory, HOMO-LUMO energy, Python parsing, Gaussian, Vibrational frequency, Basis set

1. Introduction

Computational chemistry has become an essential component of modern chemical research by providing theoretical tools that allow scientists to investigate molecular systems at the atomic and electronic level[1]. By applying principles of quantum mechanics and numerical algorithms, computational methods enable the prediction of molecular structures, energies, and physical properties without the need for laboratory experimentation. These theoretical approaches complement experimental chemistry by allowing researchers to explore systems that may be difficult to study experimentally due to cost, safety concerns, or experimental limitations[2]. As computing power and software capabilities have advanced, computational chemistry has become widely used in areas such as drug design, materials science, environmental chemistry, and molecular physics. Among the various theoretical approaches available in computational chemistry, Density Functional Theory (DFT) has become one of the most widely used methods for studying molecular electronic structure. The foundations of DFT

were established by Pierre Hohenberg and Walter Kohn in 1964, who demonstrated that the ground-state properties of a many-electron system can be uniquely determined by its electron density rather than the many-body wavefunction[3]. Because electron density depends only on three spatial coordinates, whereas the wavefunction depends on $3N$ coordinates for a system containing N electrons, the use of electron density greatly simplifies quantum mechanical calculations[4]. This reduction in computational complexity allows DFT to provide a balance between computational efficiency and predictive accuracy, making it particularly suitable for studying molecules of moderate size.

Using DFT, a variety of molecular properties can be calculated, including optimized molecular geometries, total electronic energies, vibrational frequencies, and electronic orbital energies[5]. Of particular importance in molecular electronic structure analysis are the energies of the Highest Occupied Molecular Orbital (HOMO) and the Lowest Unoccupied Molecular Orbital (LUMO)[6]. These orbitals play a key role in determining the chemical reactivity, stability, and electronic behavior of molecules. The energy difference between these orbitals, commonly referred to as the HOMO-LUMO gap, is frequently used as an indicator of molecular stability and the likelihood of electronic transitions. In quantum chemical calculations, the accuracy of predicted molecular properties depends strongly on the theoretical approximations used in the calculation. Two of the most important factors influencing computational results are the choice of computational method and the choice of basis set[7]. The computational method determines how electron-electron interactions are approximated in the system. For example, Hartree-Fock theory approximates electron interactions through an averaged electrostatic field but neglects electron correlation effects, which can lead to deviations in predicted energies. DFT addresses this limitation by incorporating electron correlation through exchange-correlation functionals, which approximate the effects of electron interactions. Different functionals, such as B3LYP, PBE, and M06, employ different mathematical formulations, which can influence the predicted molecular energies and properties.

In addition to the computational method, quantum chemical calculations require the use of a basis set. A basis set is a collection of mathematical functions used to approximate the atomic orbitals of electrons within a molecule[8]. Since exact solutions to the Schrödinger equation are not available for most molecular systems, these basis functions provide a practical approximation for describing electron distributions. Smaller basis sets require fewer functions and therefore reduce computational cost, but they may also limit the accuracy of the calculated results. Larger and more flexible basis sets allow a more accurate representation of electron density and molecular bonding. Many basis sets also include polarization functions, which allow orbitals to change shape during chemical bonding, and diffuse functions, which improve the description of electrons that occupy more spatially extended regions of space. Consequently, the choice of basis set can significantly affect calculated total energies, orbital energies, and vibrational properties.

Modern computational chemistry software packages, such as Gaussian 09, generate extensive output files that contain detailed information about molecular structures, energies,

vibrational modes, and electronic orbitals. While these outputs provide valuable information, extracting relevant data manually can be time-consuming and inefficient. Integrating computational chemistry with programming tools offers an effective solution to this challenge. Programming languages such as Python allow automated extraction and organization of relevant chemical information from Gaussian output files, enabling efficient analysis of computational results. This article aims to demonstrate how computer science techniques can be applied to support computational chemistry workflows. Molecular simulations are performed using Gaussian 09 to obtain optimized geometries, electronic energies, vibrational frequencies, and molecular orbital energies for selected small molecules. The molecules studied for the research are, (1) Carbon dioxide, (2) Methane, (3) Ethyne, and (4) Hydrogen Cyanide. Python scripts are then used to extract relevant information from Gaussian log files and organize the data into structured datasets for analysis and visualization. This approach illustrates how programming and computational chemistry can be integrated to improve the efficiency of molecular data processing and to facilitate clearer interpretation of theoretical chemical results.

2. Computational Methodology

2.1 Quantum Chemical Calculations

All quantum chemical calculations were performed using the Gaussian 09 computational chemistry software package. Molecular structures for the selected compounds were first constructed using molecular modeling software and then subjected to geometry optimization using Density Functional Theory. Geometry optimization was carried out to determine the lowest-energy structure corresponding to the equilibrium configuration of each molecule. The electronic structure calculations of (1) Carbon dioxide, (2) Methane, (3) Ethyne, and (4) Water were performed using the B3LYP exchange-correlation functional, which is widely used in computational chemistry due to its balance between computational efficiency and accuracy for organic molecules. The atomic orbitals were represented using the 6-31G basis set, which includes polarization functions that improve the description of electron distributions during chemical bonding.

Following geometry optimization, vibrational frequency calculations were performed to verify that the optimized structures correspond to true energy minima on the potential energy surface. The absence of imaginary frequencies confirms that the optimized geometries represent stable molecular configurations. Gaussian output files were analyzed to extract several key molecular properties, including total electronic energies, HOMO and LUMO orbital energies, and vibrational frequency information. These values provide insight into molecular stability, bonding characteristics, and potential chemical reactivity. To facilitate analysis of the computational results, Python scripts were developed to parse Gaussian log files and extract relevant numerical data. The scripts identify specific sections within the output files containing orbital energies and vibrational frequencies, and then organize these data into structured datasets. The processed data were subsequently used to generate graphical representations such as infrared

spectra and molecular orbital energy diagrams. This integration of Gaussian simulations with Python-based data processing provides a semi-automated workflow for organizing computational chemistry results, improving both efficiency and reproducibility.

2.2 Prototype Data Extraction Script

To simplify the analysis of Gaussian output files, a prototype parser was developed using Python. The script reads Gaussian log files and extracts selected molecular descriptors using pattern recognition techniques (**Figure 1**). The parameters extracted include: (i) total electronic energy (SCF energy), (ii) vibrational frequencies, (iii) Highest Occupied Molecular Orbital (HOMO) energy, and (iv) Lowest Unoccupied Molecular Orbital (LUMO) energy. Gaussian output files contain extensive text-based information in which relevant numerical values are embedded within specific sections. The script identifies these sections by searching for characteristic keywords, such as “SCF Done” for electronic energies and “eigenvalues” for molecular orbital energies. Vibrational frequencies are extracted from sections labeled “Frequencies –”. Once identified, the corresponding numerical values are parsed using string matching and regular expression techniques and stored in variables for further analysis. The extracted values are used directly for analysis and visualization, enabling efficient interpretation of computational results without the need for manual data transcription (**Figure 2-5**). Although the present implementation is a simple prototype, it demonstrates how basic programming tools can automate the extraction of quantum chemical results and reduce the time required for data analysis.

```
import re
import os

def extract_gaussian_energy(log_file):
    pattern = re.compile(
        r'SCF Done:\s+E\([^\)]+\)\s+=\s+([-\+]?[d*]\.[d+])',
        re.IGNORECASE
    )

    energy = None

    with open("test.log", "r") as f:
        for line in f:
            match = pattern.search(line)
            if match:
                energy = float(match.group(1))

    return energy

logfile = "test.log"

energy = extract_gaussian_energy(logfile)

filename = os.path.basename(logfile)

if energy is not None:
    print(f"File: {filename}")
    print(f"Final SCF Energy (Hartree): {energy}")
else:
    print(f"File: {filename}")
    print("No SCF energy found.")
```

Figure 1. The screenshot of the representative Python script

To assess the reliability of the extraction procedure, a subset of Gaussian output files was manually examined and compared with the values obtained from the script. Key parameters, including total electronic energy (SCF energy), HOMO and LUMO energies, and selected vibrational frequencies, showed complete numerical agreement within the precision reported by Gaussian 09 outputs. No discrepancies were observed beyond standard floating-point rounding limits (on the order of 10^{-6} Hartree for energies), confirming that the parsing approach accurately retrieves relevant molecular descriptors.

```
C:\Users\Abhiram Pannala\OneDrive\Desktop\hw\CO2-DFT-631G> python CO2-DFT-631G.py
SCF Energies found:
Step 1: -188.500175844 Hartree
Step 2: -188.500268526 Hartree
Step 3: -188.50026967 Hartree

Final SCF Energy: -188.50026967 Hartree

C:\Users\Abhiram Pannala\OneDrive\Desktop\hw\CO2-DFT-631G>|
```

Figure 2. Screenshot illustrating the extraction of electronic energy values for CO₂ from Gaussian 09 output using a Python-based parsing script.

```
C:\Users\Abhiram Pannala\OneDrive\Desktop\hw\CH4_DFT_6-31G>python CH4_DFT_6-31G_FREQ.py
Frequencies found (cm-1):
1404.6869
1404.6869
1404.6869
1602.5597
1602.5597
3042.4113
3165.2415
3165.2415
3165.2416

Total modes found: 9

C:\Users\Abhiram Pannala\OneDrive\Desktop\hw\CH4_DFT_6-31G>|
```

Figure 3. Screenshot illustrating the extraction of frequency values for CH₄ from Gaussian 09 output using a Python-based parsing script.

```

C:\Users\Abhiram Pannala\OneDrive\Desktop\hw\C2H2-DFT-6-31G>python C2H2_GEOMETRY.py
Final Standard Orientation (Angstrom):

(6, 0.0, 0.0, 0.601822)
(6, 0.0, 0.0, -0.601822)
(1, 0.0, 0.0, 1.664048)
(1, 0.0, 0.0, -1.664048)

Saved as final_geometry.xyz

C:\Users\Abhiram Pannala\OneDrive\Desktop\hw\C2H2-DFT-6-31G>

```

Figure 4. Screenshot illustrating the extraction of geometry for C₂H₂ from Gaussian 09 output using a Python-based parsing script.

```

C:\Users\Abhiram Pannala\OneDrive\Desktop\hw\C2H2-DFT-6-31G>python C2H2_DFT_6-31G_HOMO.py

Alpha Spin:
HOMO Energy: -0.28669 Hartree
LUMO Energy: 0.05256 Hartree
HOMO-LUMO Gap: 0.33925 Hartree
HOMO-LUMO Gap: 9.23146745 eV

Closed-shell system detected (RB3LYP/RHF).

C:\Users\Abhiram Pannala\OneDrive\Desktop\hw\C2H2-DFT-6-31G>

```

Figure 5. Screenshot illustrating the extraction of HOMO-LUMO values for C₂H₂ from Gaussian 09 output using a Python-based parsing script.

2.3 Data Organization and Visualization

After extraction, the molecular descriptors were organized into structured datasets for analysis. The processed data were then used to generate visual representations of molecular properties. Two types of visualizations were produced: (1) Infrared spectra derived from vibrational frequency data, (2) Frontier molecular orbital energy diagrams illustrating HOMO-LUMO gaps. These visualizations help translate raw computational output into interpretable scientific information.

3. Results and Discussion

3.1 Geometry Optimization Results

Geometry optimization calculations produced stable molecular structures for all molecules studied (**Figure 6**, **Table 1**). The absence of imaginary vibrational frequencies confirmed that the optimized geometries correspond to local minima on the potential energy surface. The optimized bond lengths and angles were consistent with experimentally reported values for these molecules, confirming the reliability of the chosen DFT method.

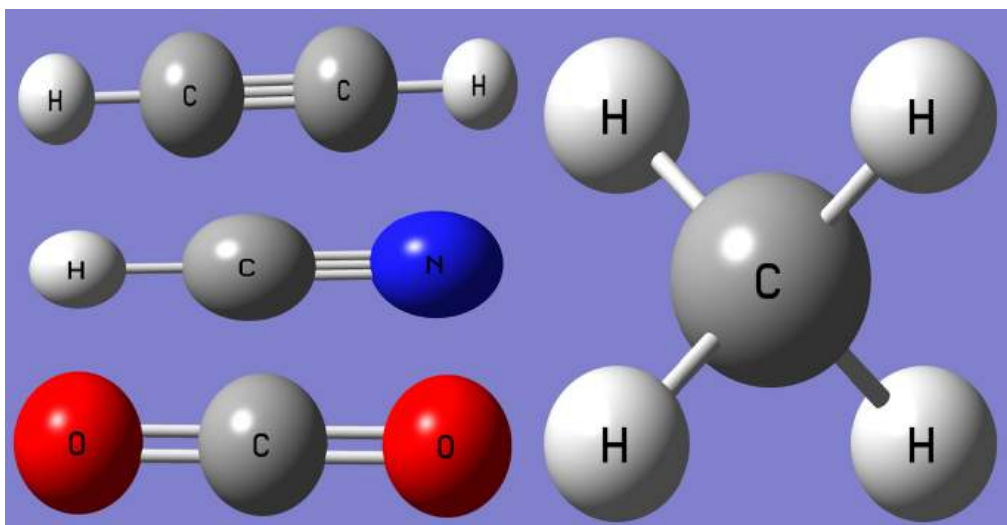


Figure 6. Optimised structures of the molecules (B3LYP/6-31G)

Geometry optimization calculations produced stable molecular structures for all molecules studied (**Table 1**). The absence of imaginary vibrational frequencies confirmed that the optimized geometries correspond to local minima on the potential energy surface. The optimized bond lengths and angles were consistent with experimentally reported values for these molecules, confirming the reliability of the chosen DFT method.

Table 1. Table depicting the zero-point energies of the molecules

Sl. No.	Molecule	Zero-point energy (Hartree)
1	CH ₄	-40.180141
2	CO ₂	-188.50027
3	HCN	-93.392488
4	C ₂ H ₂	-77.31189

3.2 Vibrational Frequency Analysis

Vibrational frequency calculations provide valuable information about molecular bonding and dynamic behavior. The prototype Python parser successfully extracted vibrational frequencies from Gaussian output files and organized them for visualization. Using the extracted frequencies and infrared intensities, simulated infrared spectra were generated (**Table 2**). Characteristic vibrational modes corresponding to functional groups were clearly observed. For example, strong O-H stretching vibrations were observed in molecules containing hydroxyl groups, such as methanol and ethanol[9]. Carbonyl stretching vibrations were also detected in molecules containing C=O bonds, including acetone and formaldehyde. These results

demonstrate that automated extraction of vibrational data can facilitate rapid generation of spectroscopic plots.

Table 2. The table of vibrational frequencies extracted from the Gaussian output file through Python scripts.

Sl. No.	Molecule	Vibrational Frequencies (cm ⁻¹)
1	CH ₄	1404.32 (td), 1602.43 (dd), 3044.23 (td), 3167.24 (td)
2	CO ₂	563.00 (dd), 1282.45, 2283.13
3	HCN	778.45 (dd), 2148.95, 3498.68
4	C ₂ H ₂	652.52 (dd), 787.49 (dd), 2070.89, 3449.94, 3557.98

The vibrational frequency analysis reveals the presence of both doubly degenerate (dd) and triply degenerate (td) modes, which arise from the symmetry characteristics of the molecules. In highly symmetric systems such as methane, triply degenerate (td) modes are observed due to its tetrahedral geometry, where equivalent vibrational motions occur along three mutually perpendicular directions with identical energy. Doubly degenerate (dd) modes are also present, corresponding to vibrational motions that occur in two equivalent planes or directions. For linear molecules such as carbon dioxide, hydrogen cyanide, and ethyne, degeneracy is primarily associated with bending vibrations. These bending modes occur in two perpendicular directions relative to the molecular axis, leading to doubly degenerate (dd) vibrations. The presence of such degeneracies is a direct consequence of molecular symmetry and confirms the expected theoretical behavior of both linear and non-linear systems. Overall, the identification of dd and td modes in the computed spectra supports the accuracy of the vibrational analysis and is consistent with established principles of molecular symmetry and normal mode theory.

3.3 Frontier Molecular Orbital Analysis

Frontier molecular orbitals play an important role in determining chemical reactivity and electronic properties. The energies of the HOMO and LUMO were extracted using the Python script. The energy difference between these orbitals, known as the HOMO-LUMO gap, provides insight into molecular stability and reactivity (**Table 3**). Molecules with smaller HOMO-LUMO gaps generally exhibit higher chemical reactivity, while larger gaps are associated with greater electronic stability. The automated extraction process allowed rapid comparison of orbital energies across the set of molecules studied.

Table 3. The table of HOMO-LUMO energies extracted from the Gaussian output file through Python scripts.

Sl. No.	Molecule	E _{LUMO} (Hartree)	E _{HOMO} (Hartree)	ΔE (Hartree)
1	CH ₄	0.11726	-0.38987	0.50713
2	CO ₂	-0.00350	-0.37293	0.36943
3	HCN	0.01208	-0.36609	0.37817
4	C ₂ H ₂	0.05040	-0.28550	0.3359

3.4 Evaluation of the Prototype Automation Approach

The prototype Python script significantly reduced the effort required to extract key molecular parameters from Gaussian output files. Instead of manually searching through large log files, the script automatically identified relevant sections and extracted numerical values within seconds. Although the present workflow was applied to a small set of molecules, the approach demonstrates the potential of automation in computational chemistry data analysis. Even simple parsing tools can improve efficiency and minimize the risk of transcription errors.

4. Educational and Research Implications

The integration of computational chemistry and programming provides valuable opportunities for interdisciplinary learning. Students who engage with such workflows gain experience in both molecular modeling and scientific data processing. The approach presented in this work illustrates how programming skills can enhance the interpretation of quantum chemical calculations. By automating routine tasks such as data extraction and plotting, researchers can focus more effectively on the scientific interpretation of results. Such workflows are particularly useful in educational settings where students are introduced to computational chemistry for the first time. The present workflow represents a prototype implementation designed to demonstrate the feasibility of automated data extraction from Gaussian output files. The parser currently focuses on a limited set of molecular descriptors and may require modification to support additional properties or other computational chemistry software packages. Future developments may include expanding the parser to extract thermodynamic data, reaction energies, and additional electronic structure parameters. Integration with machine learning techniques could also enable predictive modeling of molecular properties based on computational datasets.

5. Limitations of the Study

While the present study demonstrates the feasibility of integrating quantum chemical calculations with Python-based data extraction, several limitations should be acknowledged.

First, the implemented parser is designed specifically for the output format of Gaussian 09 and relies on fixed text patterns for identifying molecular properties. As a result, variations in output formatting, differences in calculation types, or the use of other computational chemistry software packages may require modification of the parsing logic. The current implementation extracts a limited set of molecular descriptors, namely total electronic energies, vibrational frequencies, and frontier molecular orbital energies. Other important properties, such as thermodynamic corrections, charge distributions, and excited-state energies, were not included in the present workflow. Expanding the scope of extracted parameters would provide a more comprehensive analysis of molecular systems. The workflow operates on individual Gaussian output files and does not yet implement fully automated batch processing or large-scale dataset generation. Although the prototype demonstrates the potential for automation, further development is required to enable high-throughput analysis of extensive molecular libraries. Additionally, the accuracy of the computational results is inherently dependent on the choice of theoretical method and basis set. In this study, a single level of theory was employed, and no systematic comparison with alternative methods or experimental data was performed. Consequently, the results should be interpreted within the context of the selected computational approach. Finally, the Python script currently performs direct extraction and basic analysis without incorporating advanced error handling for incomplete or failed calculations. Future improvements could include validation routines to detect convergence failures, missing data sections, or inconsistencies in output files. Despite these limitations, the study provides a proof-of-concept framework demonstrating how programming tools can be integrated with computational chemistry workflows to improve efficiency and data handling.

6. Conclusion

This study demonstrates a prototype workflow integrating quantum chemical calculations with automated data extraction using Python. Density Functional Theory calculations were performed using Gaussian 09 to obtain optimized geometries, vibrational frequencies, and molecular orbital energies for a set of small molecules. A simple Python script was developed to parse Gaussian output files and extract selected molecular descriptors. The extracted data were organized into structured datasets and used to generate visual representations such as infrared spectra and HOMO-LUMO energy diagrams. Although the present implementation represents a small-scale demonstration, it highlights the potential of automation tools for improving efficiency in computational chemistry research. The integration of programming and theoretical chemistry provides a practical pathway for managing computational data and enhancing scientific analysis.

Works Cited

- [1]J. A. Keith et al., “Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems,” *Chem. Rev.*, vol. 121, no. 16, pp. 9816–9872, Aug. 2021, doi: 10.1021/acs.chemrev.1c00107.
- [2]G.-J. Cheng, X. Zhang, L. W. Chung, L. Xu, and Y.-D. Wu, “Computational Organic Chemistry: Bridging Theory and Experiment in Establishing the Mechanisms of Chemical Reactions,” *J. Am. Chem. Soc.*, vol. 137, no. 5, pp. 1706–1725, Feb. 2015, doi: 10.1021/ja5112749.
- [3]P. Hohenberg and W. Kohn, “Inhomogeneous Electron Gas,” *Physical Review*, vol. 136, no. 3B, pp. B864–B871, Nov. 1964, doi: 10.1103/PhysRev.136.B864.
- [4]E. J. Baerends and O. V. Gritsenko, “A Quantum Chemical View of Density Functional Theory,” *J. Phys. Chem. A*, vol. 101, no. 30, pp. 5383–5403, Jul. 1997, doi: 10.1021/jp9703768.
- [5]A. A. El-Azhary and H. U. Suter, “Comparison between Optimized Geometries and Vibrational Frequencies Calculated by the DFT Methods,” *J. Phys. Chem.*, vol. 100, no. 37, pp. 15056–15063, Jan. 1996, doi: 10.1021/jp960618o.
- [6]B. Ganiev, U. Mardonov, and G. Kholikova, “Molecular structure, HOMO-LUMO, MEP - – Analysis of triazine compounds using DFT (B3LYP) calculations,” *Mater. Today Proc.*, Oct. 2023, doi: 10.1016/j.matpr.2023.09.191.
- [7]A. D. Boese, J. M. L. Martin, and N. C. Handy, “The role of the basis set: Assessing density functional theory,” *J. Chem. Phys.*, vol. 119, no. 6, pp. 3005–3014, Aug. 2003, doi: 10.1063/1.1589004.
- [8]A. D. Boese, J. M. L. Martin, and N. C. Handy, “The role of the basis set: Assessing density functional theory,” *J. Chem. Phys.*, vol. 119, no. 6, pp. 3005–3014, Aug. 2003, doi: 10.1063/1.1589004.
- [9]S. Burikov, T. Dolenko, S. Patsaeva, Y. Starokurov, and V. Yuzhakov, “Raman and IR spectroscopy research on hydrogen bonding in water–ethanol systems,” *Mol. Phys.*, vol. 108, no. 18, pp. 2427–2436, Sep. 2010, doi: 10.1080/00268976.2010.516277.

Determinants of Psychological Distress in Long-Term Cancer Survivors By Joshua Han

Abstract

Cancer survivorship has improved over recent decades. However, many long-term survivors continue to experience persistent depression and anxiety. Although the mental health of cancer survivors has received increasing attention, psychological research remains limited compared to studies on cancer diagnosis and treatment. To address this gap, this study examined psychological distress in long-term cancer survivors compared to healthy controls and identified vulnerable demographic and national level factors related with mental health outcomes. The results indicate that younger survivors are particularly susceptible to anxiety. Moreover, high national life satisfaction is suggested to be a predictor of psychological resilience among cancer survivors. These findings suggest that cancer recovery should be understood not only in terms of physical outcomes but also in relation to psychological well-being and social reintegration.

Introduction

Approximately 10 million deaths were attributed to cancer worldwide in 2022, making cancer the second leading cause of death (Langselius et al., 2026; Tran et al., 2022). The global burden of cancer is increasing, with projections indicating that nearly 20 million individuals in the United States will have a history of cancer by 2030 and about 50 million survivors globally (Mariotto et al., 2011). Over recent decades, survival rates have improved substantially due to advances in early detection and targeted treatments, with nearly 70% of patients now surviving at least five years after diagnosis (Mitchell et al., 2013). While this is widely regarded as a major achievement, physical recovery does not align with the mental health outcomes of long-term cancer survivors (Breidenbach et al., 2022; Endo et al., 2022). For example, Zarei et al. (2025) demonstrated that difficulties in daily activities and changes in social roles can hinder survivors from returning to their normal lives.

Many long-term cancer survivors continue to experience psychological challenges after being declared cancer-free, such as fear of cancer recurrence and emotional distress related to having faced a life-threatening disease (Luigjes-Huizer et al., 2022). Psychological distress in cancer survivors often appears as symptoms of depression and anxiety (Endo et al., 2022). Furthermore, Walker et al. (2020) suggested that depressive symptoms (e.g., deep sadness, fatigue, changes in appetite) and anxiety symptoms (e.g., restlessness, dizziness, persistent tension) often occur simultaneously among survivors. Since depression and anxiety can greatly reduce quality of life and may lead to difficulties in daily functioning, these mental health problems may result in serious consequences (Endo et al., 2022; Akechi et al., 2010). In severe cases, untreated mood disorders can be associated with a higher risk of suicidal thoughts and even mortality, highlighting the importance of mental health care in cancer survivorship (Akechi et al., 2010).

Despite the growing attention to cancer survivorship, psychological research focusing on long-term cancer survivors is scarce relative to the stages of cancer diagnosis and acute treatment

(Breidenbach et al., 2022; Carreira et al., 2018). Moreover, prior studies highlight the need for additional research into mental health problems among survivors, particularly in identifying risk factors (Breidenbach et al., 2022; Carreira et al., 2018; Syrowatka et al., 2017). Importantly, Wang et al. (2018) suggested that depression and anxiety in cancer survivors may serve as predictors of cancer recurrence and survival outcomes. To address the knowledge gap, this study aims to measure the psychological burden of long-term cancer survivors and identify the socioeconomic factors that affect their mental well-being. Specifically, this study synthesizes evidence from peer-reviewed studies using standardized scales, including the Center for Epidemiologic Studies Depression Scale (CES-D), Patient Health Questionnaire-9 (PHQ), and Hospital Anxiety and Depression Scale (HADS). Based on this approach, the objective of this study is to identify vulnerable subgroups among cancer survivors who are at increased risk for mood disorders, thereby supporting more targeted diagnosis and intervention.

Methods

Literature Search and Data Selection

A systematic literature search was conducted to investigate the psychological distress of cancer survivors. Relevant literature reporting the distribution of standardized psychological scales (CES-D, PHQ, HADS-D, and HADS-A) was selected through Google Scholar and PubMed. The study had to include the parameters of the number of participants (N), mean (μ), and standard deviation (σ). A total of 16 studies met these criteria and were included in the analysis (Götze et al., 2022; Endo et al., 2022; Breidenbach et al., 2022; Tran et al., 2021; Stagl et al., 2015; Reyes-Gibby et al., 2012; Hinz et al., 2012; Brunault et al., 2012; Kim et al., 2010; Boyes et al., 2009; Crespi et al., 2008; Bishop et al., 2007; van Wilgen et al., 2006; Dahl et al., 2005; Andrykowski et al., 2005; Fosså et al., 2003). Moreover, the control data from the 16 selected studies were utilized to establish a baseline for the general population. This ensured that the healthy control comparisons were as consistent as possible with the methodology used for survivor groups.

To identify which sociodemographic subgroups exhibit a higher prevalence of mood disorders, age and geographic location were set as primary variables. Moreover, to analyze the correlation between national prevalence of mood disorders and external environmental factors, two global indices were incorporated from 2019: the Index of Cancer Preparedness (ICP) (The Economist Intelligence Unit, 2019) and the National Happiness Index (Helliwell et al., 2019). ICP indicates the country's medical infrastructure and readiness for cancer and the national happiness index represents overall life satisfaction and social well-being. Data from 2019 were selected to reflect the pre-pandemic period and avoid potential psychological confounding effects of COVID-19.

Statistical Analysis

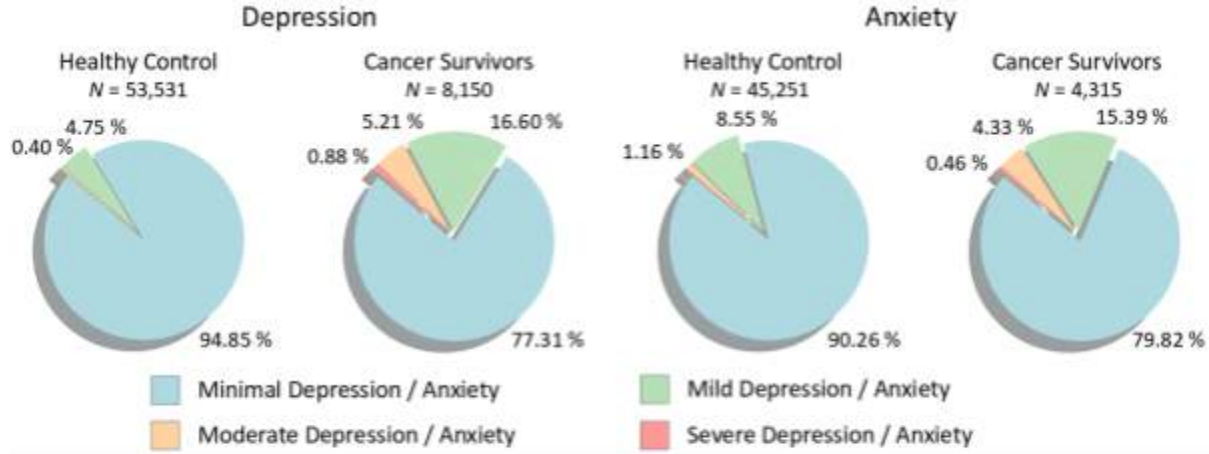


Fig. 1: Comparison of mental health prevalence between healthy controls and long-term cancer survivors. The pie charts illustrate the distribution of depression (left) and anxiety (right) severity levels.

Given that each survey involved at least 100 participants, with most exceeding 1,000, the psychological scores across the cohorts were assumed to be the Gaussian distribution. Based on the clinical cut-off scores for each questionnaire, the symptoms were categorized into four severity levels: minimal, mild, moderate, and severe (Breidenbach et al., 2022; Götze et al., 2019; Zich et al., 1990). Population proportions across severity categories were estimated using reported means and standard deviations from the surveys. To aggregate data across different studies using the same standardized psychological scale, the cumulative mean (μ_{total}) and standard deviation (σ_{total}) were calculated using the following pooled variance formulas (Cochran 1977):

$$\mu_{total} = \frac{\sum_{i=1}^k (n_i \cdot \mu_i)}{\sum_{i=1}^k (n_i)}$$

$$\sigma_{total}^2 = \frac{\sum_{i=1}^k [n_i \cdot (\sigma_i^2 + (\mu_i - \mu_{total})^2)]}{\sum_{i=1}^k (n_i)}$$

in which n_i is the number of participants in the i^{th} study, μ_i and σ_i represents the mean and the standard deviation of the scores in the i^{th} study, and k is the total number of individual studies.

To compare distributions and determine statistical significance, Welch's t-test were performed using the Scikit-learn (v1.8.0) library in Python (v3.8.8). Additionally, all graphical representations in the Results section were generated using the Seaborn (v0.13.2) library in Python.

Results

Comparison of Mental Health: Healthy Controls vs. Cancer Survivors

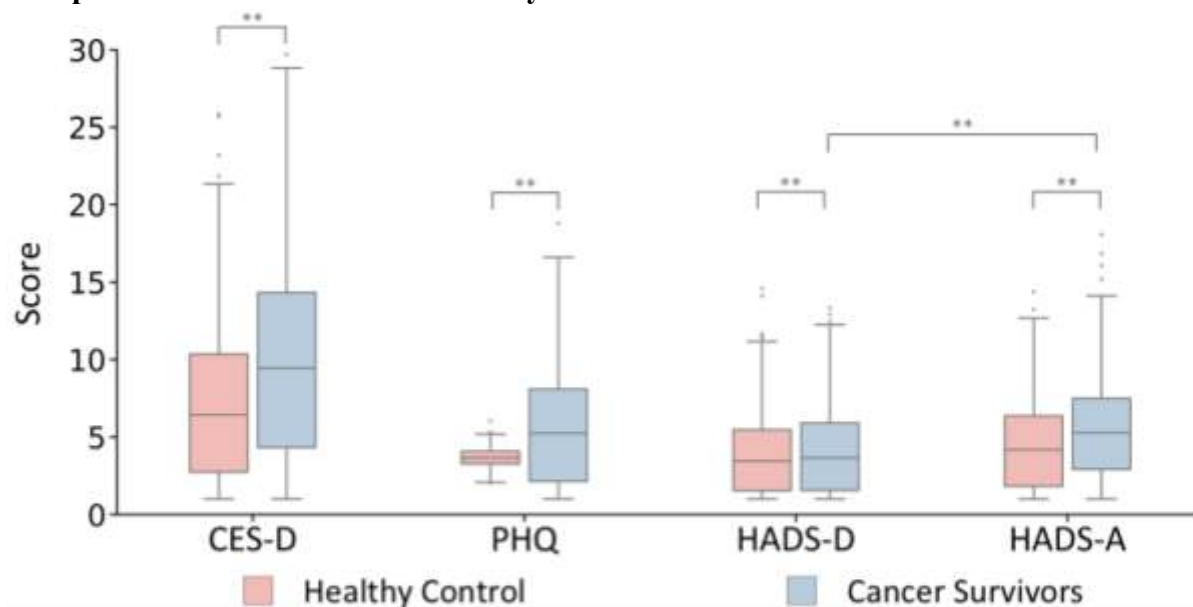


Fig. 2: Statistical disparity in psychological distress scores across standardized scales. Box plots are displayed to visually compare the distribution of scores between healthy controls (red) and cancer survivors (blue) for CES-D, PHQ, HADS-D, and HADS-A. Statistical significance was assessed using the Welch's t-test, with labeling a double asterisk (**) if the p value was less than 0.005.

Analysis of psychological distress revealed a significant disparity between healthy controls and long-term cancer survivors (Fig. 1). In the depression analysis ($N = 53,531$ for controls; $N = 8,150$ for survivors), the proportion of individuals with minimal symptoms dropped from 94.85% in controls to 77.31% in survivors. Similarly, in the anxiety analysis ($N = 45,251$ for controls; $N = 4,315$ for survivors), survivors showed a higher prevalence of mild-to-severe symptoms at 20.18%, compared to 9.74% in the control group. Furthermore, survivors exhibited more severe distress across all four standardized scales (Welch's t-test, $p < 0.005$; Fig. 2). Specifically mean ($\pm$ standard deviation) scores for survivors increased notably compared to controls: CES-D rose from 6.22 ± 6.08 to 9.20 ± 7.17 , PHQ from 3.65 ± 0.54 to 4.90 ± 4.39 , HADS-D from 3.44 ± 2.92 to 3.89 ± 3.25 , and HADS-A from 4.02 ± 3.09 to 5.07 ± 3.41 . Also, comparative analysis between the HADS subscales indicated that survivors experience higher levels of anxiety relative to depressive symptoms (Welch's t-test, $p < 0.005$).

Age-related Vulnerability in Long-term Cancer Survivors

To analyze relative vulnerability within the long-term cancer survivor population, the relationship between age and psychological scores was examined using the age group with the lowest mean score as a baseline (Fig. 3). A negative correlation was observed between age and anxiety, with a linear regression coefficient (β) of -0.0391 and a Pearson Correlation coefficient (R^2) of 0.615. This indicates that younger survivors are more vulnerable to anxiety. For example,

data from Boyes et al. (2009) observed that survivors in their thirties faced anxiety levels nearly four times higher than those in their seventies. While depression followed a similar downward trajectory, the results across the five analyzed datasets (Breidenbach et al., 2022; Endo et al., 2022; Götze et al., 2019; Boyes et al., 2009; Fosså et al., 2003) were more varied for depression ($R^2 = 0.053$) than for anxiety (Fig. 3).

Global Prevalence and Socioeconomic Correlations

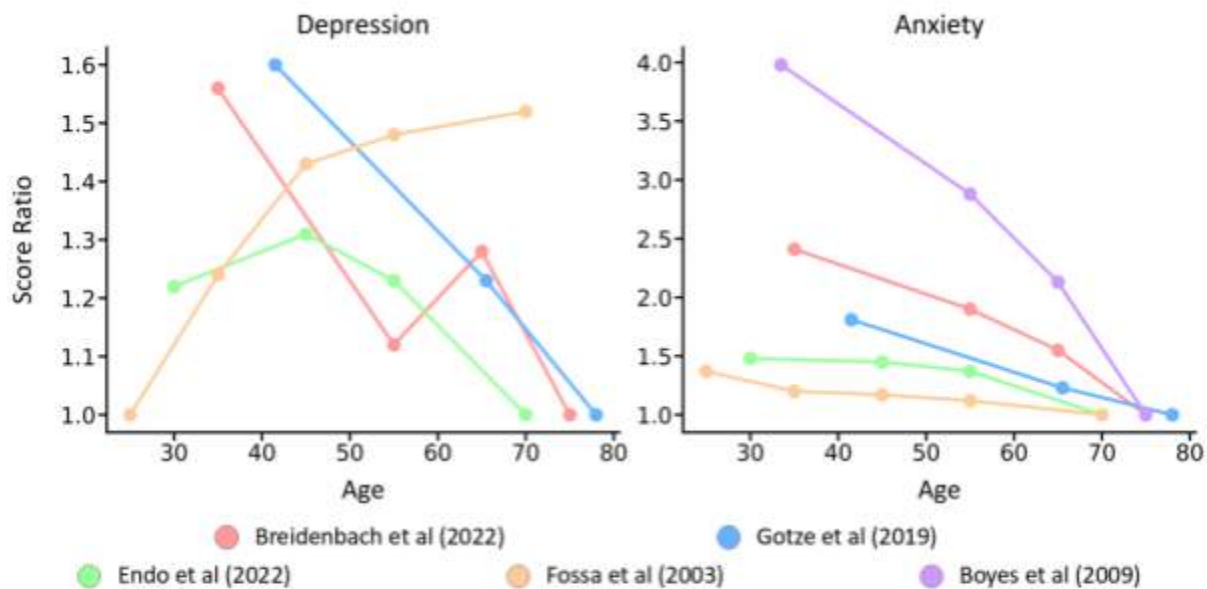


Fig. 3: Vulnerability in depression and anxiety among cancer survivors in terms of age. The relative score ratios for depression (left) and anxiety (right) were plotted against age, using the lowest mean score group as a baseline.

The national environment was found to influence survivors' mental health, with wide disparities in resilience across several countries (Fig. 4A). Germany reported the highest distress, with only 48.54% of survivors reporting minimal depression and 60.70% reporting minimal anxiety. In contrast, Norway exhibited the highest resilience, with 95.18% of survivors reporting minimal symptoms. To identify driving factors, linear regression models were used (Fig. 4B). Germany was identified as a significant outlier and excluded from these models to prevent bias. Medical readiness, measured by the ICP, showed no correlation with depression or anxiety ($\beta = 0.02$, $R^2 < 0.01$). Conversely, a much stronger negative correlation was found with the National Happiness Index ($\beta = -3.79$, $R^2 = 0.22$), suggesting that countries with higher overall life satisfaction and social support systems maintain lower rates of psychiatric disorders among survivors.

Discussion

In this study, a systematic analysis was conducted to investigate the psychological distress among cancer survivors and the factors contributing to their vulnerability. Results

showed that long-term cancer survivors experience significantly higher levels of mental health problems across all measured scales (CES-D, PHQ, HADS-D, and HADS-A) compared to healthy controls (Fig. 1, Fig. 2). Moreover, Fig. 2 indicates that anxiety is a more prevalent and severe issue than depression among these survivors. This difference may be explained by prior findings. Oh & Cho (2022) suggested that depression is often associated with physical pain and fatigue, which tend to diminish over time as physical recovery progresses. In contrast, anxiety is primarily driven by the fear of cancer recurrence, representing a persistent concern about the future (Simard et al., 2013).

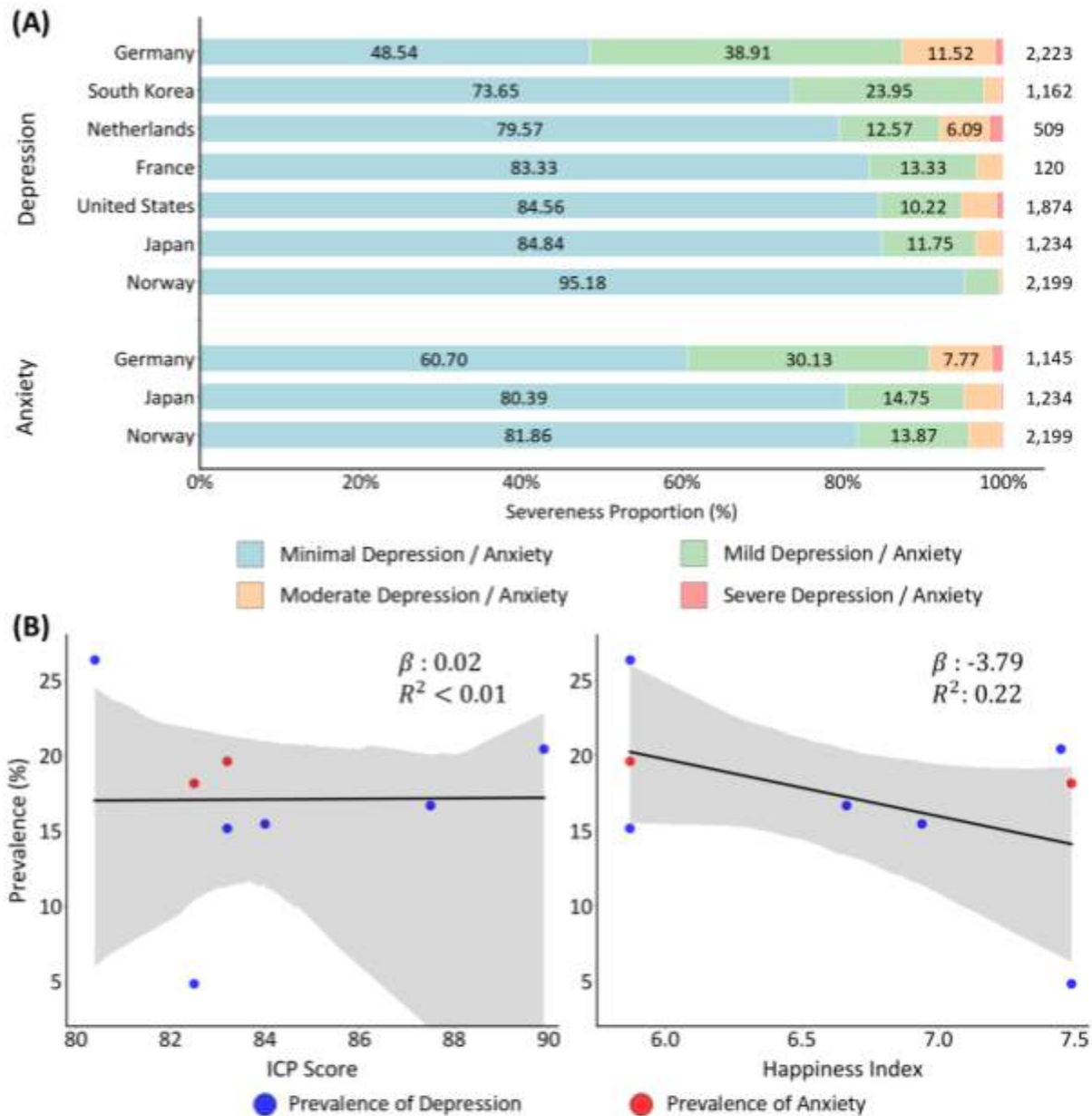


Fig. 4: Global variations in prevalence and socioeconomic factors. (A) Bar graph showing the disparity in psychological resilience among long-term cancer survivors among countries. Numbers displayed on the right refer to the number of participants by country. (B) Correlation of depression and anxiety prevalence with the Index of Cancer Preparedness (left) and National Happiness Index (right). Gray shades indicate the 95% confidence intervals regarding the black regression line.

When examining which groups are most vulnerable to psychological distress, age emerged as a critical factor. As shown in Fig. 3, younger cancer survivors experienced higher

levels of anxiety than older survivors, whereas no significant association was observed between age and depression. This finding is consistent with previous studies that were not included in the systematic analysis (Weiss Wiesel et al., 2015; Hinz et al., 2017; Hinz et al., 2009). This difference may be attributed to the substantial life disruptions cancer causes during early adulthood. For younger individuals, a cancer diagnosis can interrupt key life milestones, such as career development, financial independence, and family formation (Bellizzi et al., 2012). This disruption often leads to feelings of isolation, financial strain, and a sense of being left behind compared to their healthy peers (Zebrack 2011). In contrast, older survivors tend to demonstrate greater psychological resilience, as they often evaluate their well-being in comparison to their peers (Götze et al., 2019). Given that aging is commonly accompanied by chronic illness and physical decline, older survivors may perceive their cancer history as a part of the normal aging process rather than a uniquely adverse event (Lo et al., 2010). This shift in perspective may help buffer older survivors from the heightened anxiety observed in younger patients.

Beyond individual factors, this study examined how the national environment influences the mental health of cancer survivors. As shown in Fig. 4B, the Index of Cancer Preparedness, which measures the quality of clinical facilities and medical readiness, was not associated with rates of depression and anxiety among survivors. In contrast, a negative correlation was observed with the National Happiness Index, indicating that countries with higher overall life satisfaction tend to report lower rates of psychological distress among survivors. This finding highlights a potential gap in survivorship care. While advanced medical facilities and treatments are essential for physical recovery, they may be insufficient for long-term psychological well-being. After the completion of active treatment, survivors' mental health is likely influenced by psychosocial factors such as community support, financial stability, and social connectedness (Bellizzi et al., 2012). For instance, a nationwide study in South Korea demonstrated that strong social support systems serve as a critical protective buffer against emotional distress related to cancer, reducing the risk of depression and anxiety (Eom et al., 2013). Furthermore, Zebrack 2000 emphasized that the transition back to everyday life requires a supportive societal environment rather than routine clinical monitoring alone. Therefore, these findings suggest that improving the long-term well-being of cancer survivors requires not only strong healthcare systems but also active investment in supportive community environments.

Despite the meaningful findings of this study, several limitations should be acknowledged. First, data were aggregated from 16 peer-reviewed studies to ensure a sufficiently large sample. As a result, the dataset includes a heterogeneous mix of cancer types, disease stages, and treatment modalities. Survivors were analyzed as a single group without accounting for clinical heterogeneity, which may have obscured subgroup differences in psychological outcomes. Second, psychological scores were assumed to follow a Gaussian distribution to estimate the proportion of individuals experiencing varying levels of psychological distress. Although this approach is commonly used for large samples, the estimates may not fully reflect actual score distributions. Finally, in the analysis of international trends in Fig. 4B, Germany was excluded from the linear regression model. This exclusion was necessary to identify the overall

trend, but it may suggest that indices such as the National Happiness Index do not fully capture national systemic factors influencing mental health.

In conclusion, this study suggests that achieving a cancer-free status does not necessarily indicate the end of the patient's struggle, particularly in terms of mental health. The findings show that long-term cancer survivors continue to experience a significant psychological burden, with anxiety often more prominent than depression. Moreover, younger survivors and those in environments with limited social support appear to be at higher risk of depression and anxiety. These results indicate that advanced medical care alone may be insufficient to address the long-term psychological needs of cancer survivors. Accordingly, cancer survivorship care should go beyond routine physical check-ups and include psychological and community support. Overall, this study implies that recovery from cancer should be considered in terms of both physical outcomes, psychological well-being, and social reintegration.

Works Cited

- Andrykowski, Michael A., et al. "Long-Term Health-Related Quality of Life, Health, and Spiritual Well-Being after Hematopoietic Stem-Cell Transplantation." *Journal of Clinical Oncology*, vol. 23, no. 24, 2005, pp. 599-608.
- Akechi, Tatsuo, et al. "Gender differences in factors associated with suicidal ideation in major depression among cancer patients." *Psycho-Oncology* 19.4 (2010): 384-389.
- Bellizzi, Keith M., et al. "Positive and negative psychosocial impact of being diagnosed with cancer as an adolescent or young adult." *Cancer* 118.20 (2012): 5155-5162.
- Bishop, Michelle M., et al. "Late Effects of Cancer and Hematopoietic Stem-Cell Transplantation on Spouses or Partners Compared with Survivors and Survivor-Matched Controls." *Journal of Clinical Oncology*, vol. 25, no. 11, 2007, pp. 1403-11.
- Boyes, Adrian W., et al. "Anxiety and Depression among Long-Term Survivors of Cancer in Australia: Results of a Population-Based Survey." *Medical Journal of Australia*, vol. 190, no. S7, 2009, pp. S94-S98.
- Breidenbach, C., et al. "Prevalence and Determinants of Anxiety and Depression in Long-Term Breast Cancer Survivors." *BMC Psychiatry*, vol. 22, no. 101, 2022.
- Brunault, Paul, et al. "Depression is Associated with Some Patient-Perceived Cosmetic Changes, but Not with Radiotherapy-Induced Late Toxicity, in Long-Term Breast Cancer Survivors." *Psycho-Oncology*, vol. 22, no. 3, 2013, pp. 590-97.
- Carreira, Helena, et al. "Associations between breast cancer survivorship and adverse mental health outcomes: a systematic review." *JNCI: Journal of the National Cancer Institute* 110.12 (2018): 1311-1327.
- Cochran, William G. *Sampling Techniques*. 3rd ed., Wiley, 1977.
- Crespi, Catherine M., et al. "Refinement and Psychometric Evaluation of the Impact of Cancer Scale." *Journal of the National Cancer Institute*, vol. 100, no. 21, 2008, pp. 1530-41.
- Dahl, Alv A., et al. "Study of Anxiety Disorder in Long-Term Survivors of Testicular Cancer." *Journal of Clinical Oncology*, vol. 23, no. 10, 2005, pp. 2389-95.
- Endo, M., et al. "Depressive and Anxiety Symptoms among Japanese Cancer Survivors: Japan Cancer Survivorship Research Project." *BMC Cancer*, vol. 22, no. 134, 2022.
- Eom, Chun-Sick, et al. "Impact of perceived social support on the mental health and health-related quality of life in cancer patients: results from a nationwide, multicenter survey in South Korea." *Psycho-Oncology* 22.6 (2013): 1283-1290.
- Fosså, Sophie D., et al. "Fatigue, Anxiety, and Depression in Long-Term Survivors of Testicular Cancer." *Journal of Clinical Oncology*, vol. 21, no. 7, 2003, pp. 1249-54.
- Götze, Heidi, et al. "Depression and Anxiety in Long-Term Survivors 5 and 10 Years after Cancer Diagnosis." *Supportive Care in Cancer*, vol. 28, no. 1, 2020, pp. 211-20.
- Helliwell, John F., et al. "World Happiness Report 2019." *World Happiness Report, Sustainable Development Solutions Network*, 2019, <https://data.worldhappiness.report/map>. Accessed 26 Mar. 2026.
- Hinz, Andreas, et al. "Anxiety and depression in patients with prostate cancer and other

- urogenital cancer: a longitudinal study." *Urologic oncology: seminars and original investigations*. Vol. 27. No. 4. Elsevier, 2009.
- Hinz, Andreas, et al. "Assessment of Depression Severity with the PHQ-9 in Cancer Patients and in the General Population." *BMC Psychiatry*, vol. 16, no. 22, 2016.
- Hinz, A., et al. "The relationship between global and specific components of quality of life, assessed with the EORTC QLQ-C30 in a sample of 2019 cancer patients." *European journal of cancer care* 26.2 (2017): e12416.
- Kim, S. H., et al. "Prevalence and Predictors of Anxiety and Depression Among Cervical Cancer Survivors in Korea." *International Journal of Gynecological Cancer*, vol. 20, no. 6, 2010, pp. 1017-24.
- Langselius, Oliver, et al. "Avoidable deaths through the primary prevention, early detection, and curative treatment of cancer worldwide: a population-based study." *The Lancet Global Health* 14.3 (2026): e356-e366.
- Lo, Christopher, et al. "Age and depression in patients with metastatic cancer: the protective effects of attachment security and spiritual wellbeing." *Ageing & Society* 30.2 (2010): 325-336.
- Luigjes-Huizer, Yvonne L., et al. "What is the prevalence of fear of cancer recurrence in cancer survivors and patients? A systematic review and individual participant data meta-analysis." *Psycho-Oncology* 31.6 (2022): 879-892.
- Mariotto, Angela B., et al. "Projections of the cost of cancer care in the United States: 2010–2020." *Journal of the National Cancer Institute* 103.2 (2011): 117-128.
- Oh, Pok-Ja, and Jung-Ran Cho. "Changes in fatigue, psychological distress, and quality of life after chemotherapy in women with breast cancer: a prospective study." *Cancer nursing* 43.1 (2020): E54-E60.
- Reyes-Gibby, Catherine C., et al. "Depressive Symptoms and Health-Related Quality of Life in Breast Cancer Survivors." *Journal of Women's Health*, vol. 21, no. 3, 2012, pp. 311-18.
- Simard, Sébastien, et al. "Fear of cancer recurrence in adult cancer survivors: a systematic review of quantitative studies." *Journal of cancer survivorship* 7.3 (2013): 300-322.
- Stagl, Jennifer M., et al. "Randomized Controlled Trial of Cognitive Behavioral Stress Management in Breast Cancer: A Brief Report of Effects on 5-Year Depressive Symptoms." *Health Psychology*, vol. 34, no. 2, 2015, pp. 176-80.
- Syrowatka, Ania, et al. "Predictors of distress in female breast cancer survivors: a systematic review." *Breast cancer research and treatment* 165.2 (2017): 229-245.
- The Economist Intelligence Unit. *Cancer Preparedness around the World: National Readiness for a Global Epidemic*. International Cancer Control Partnership, 2019, www.iccp-portal.org/resources/cancer-preparedness-around-world-national-readiness-global-epidemic. Accessed 26 Mar. 2026.
- Tran, Khanh Bao, et al. "The global burden of cancer attributable to risk factors, 2010–19: a systematic analysis for the Global Burden of Disease Study 2019." *The Lancet* 400.10352 (2022): 563-591.

- Tran, Thi Xuan Mai, et al. "Fear of Cancer Recurrence and Its Negative Impact on Health-Related Quality of Life in Long-term Breast Cancer Survivors." *Cancer Research and Treatment*, vol. 54, no. 2, 2022, pp. 385-96.
- Van Wilgen, C. Paul, et al. "Measuring Somatic Symptoms with the CES-D to Assess Depression in Cancer Patients after Treatment: Comparison among Patients with Oral/Oropharyngeal, Gynecological, Colorectal, and Breast Cancer." *Psychosomatics*, vol. 47, no. 6, 2006, pp. 465-70.
- Walker, Jane, et al. "Different independent associations of depression and anxiety with survival in patients with cancer." *Journal of Psychosomatic Research* 138 (2020): 110218.
- Wang, Yun-He, et al. "Depression and anxiety in relation to cancer incidence and mortality: a systematic review and meta-analysis of cohort studies." *Molecular psychiatry* 25.7 (2020): 1487-1499.
- Weiss Wiesel, Talia R., et al. "The relationship between age, anxiety, and depression in older adults with cancer." *Psycho-Oncology* 24.6 (2015): 712-717.
- Zarei, Bahare, et al. "Social relationship challenges perceived by cancer patients and survivors: a qualitative meta-synthesis." *BMC Psychology* 13.1 (2025): 1195.
- Zebrack, Brad J. "Cancer survivor identity and quality of life." *Cancer practice* 8.5 (2000): 238-242.
- Zebrack, Brad J. "Psychological, social, and behavioral issues for young adults with cancer." *Cancer* 117.S10 (2011): 2289-2294.
- Zich, Jane M., et al. "Screening for Depression in Primary Care Clinics: The CES-D and the BDI." *The International Journal of Psychiatry in Medicine*, vol. 20, no. 3, 1990, pp. 259-77.

Deep Learning-Based Multi-Class Eye Disease Classification with Grad-CAM

Interpretation By Foma Molchanov

Abstract

Early detection of eye diseases is important because it can help prevent vision loss and improve treatment outcomes. However, analyzing retinal images manually takes time and requires trained specialists, which makes large scale screening difficult. In this study, a deep learning based system was developed to classify eye diseases from retinal images. Four different models, MobileNetV2, ResNet18, Efficient Net B0, and DenseNet121, were tested to identify seven conditions: age related macular degeneration, cataract, diabetic retinopathy, glaucoma, hypertension, myopia, and normal cases. To improve performance, transfer learning and data augmentation were used so that the models could learn more effectively from the available data. Among the models tested, DenseNet121 achieved the highest accuracy of 89.55 percent with an F1 score of around 0.89. To understand how the model makes its decisions, a method called Grad CAM was used. This technique highlights the important areas in the retinal image that influence the prediction. The results showed that the model focuses on relevant regions such as lesions and structural changes in the eye. Overall, the findings suggest that deep learning can be used to classify multiple eye diseases with good accuracy. This type of system could support doctors and make large scale screening more accessible, especially in areas where medical experts are limited.

1. Introduction

Eye diseases are a major global health problem and affect millions of people worldwide. If they are not detected early, they can lead to partial or complete vision loss. Conditions such as diabetic retinopathy, glaucoma, cataract, and age-related macular degeneration are among the leading causes of blindness (Marques et al. 2021; Swenor and Ehrlich 2021). Any of these diseases develop slowly and may not show clear symptoms in the early stages, which makes timely diagnosis difficult. Diseases such as diabetic retinopathy, glaucoma, cataract, and age-related macular degeneration (AMD) are among the most common causes of vision loss (Baptista et al. 2015). Early detection is important because many eye diseases can be treated or managed if identified at the right time. However, traditional diagnosis depends on trained ophthalmologists examining retinal images manually. In many parts of the world, especially in rural or developing regions, access to such experts is limited. This creates a need for automated systems that can assist in detecting eye diseases more efficiently. Recent advances in artificial intelligence have made it possible to analyze medical images automatically. In particular, deep learning has shown strong performance in image-based tasks. A type of model called a Convolutional Neural Network is commonly used for analyzing images because it can learn patterns such as edges, textures, and more complex features like lesions. These abilities make it useful for studying retinal images, where disease related changes can be subtle and difficult to identify (Rajan and Sreejith 2019). Despite these advantages, there are still some challenges.

Many existing studies focus only on a small number of diseases, which does not reflect real clinical situations where multiple conditions may be present. In addition, retinal images can vary due to lighting conditions, camera differences, and patient factors, which makes the task more complex (Rajan and Sreejith 2019). Another important issue is that deep learning models are often difficult to interpret, meaning it is not always clear how they make decisions. This lack of transparency can reduce trust in medical applications. To address these challenges, this study develops a deep learning-based system for classifying multiple eye diseases from retinal images. Seven categories are considered: age related macular degeneration, cataract, diabetic retinopathy, glaucoma, hypertension, myopia, and normal cases. This makes the task more realistic compared to simpler classification problems. Four different models are evaluated, including MobileNetV2, ResNet18, Efficient Net B0, and DenseNet121. These models were chosen because they represent different design approaches, ranging from lightweight models to deeper networks with stronger feature learning ability. Transfer learning is used to improve performance, allowing the models to learn effectively even with limited data. To better understand how the model makes predictions, a technique called Grad CAM is applied. This method highlights important regions in the retinal image that influence the model's decision, helping to check whether the model focuses on meaningful features. The main aim of this study is to develop an accurate and reliable system for multi class eye disease classification. The results show that deep learning models, especially DenseNet121, can achieve strong performance and may be useful for supporting doctors and enabling large scale screening in real world applications.

2. Methodology

This section explains how the system was developed to classify different eye diseases using retinal images. The process includes preparing the data, training the models, and evaluating their performance. The dataset used in this study consists of retinal fundus images divided into seven categories: age related macular degeneration, cataract, diabetic retinopathy, glaucoma, hypertension, myopia, and normal cases. These images contain important features such as blood vessels, the optic disc, and the macula, which help in identifying different diseases. The task is treated as a multi class classification problem, where each image is assigned to one of the seven categories. This is more challenging than simple two class problems because some diseases look very similar.

2.1 Data Preprocessing and Augmentation

Before training, all images were resized to 224 by 224 pixels so they could be used with standard deep learning models. To improve performance, data augmentation was applied to the training images. This included flipping the images horizontally, rotating them, and slightly changing their colors. These changes help the model learn to recognize patterns under different conditions, such as changes in lighting or image orientation. For validation data, only resizing and normalization were applied so that the evaluation remains fair and realistic.

2.2 Deep Learning Model Architectures

Four different deep learning models were tested in this study.

- MobileNetV2 is a lightweight model that is fast and efficient.
- ResNet18 uses special connections that help the model learn better in deeper layers.
- Efficient Net B0 balances model size and accuracy.
- DenseNet121 connects each layer to all others, which helps reuse features and improves learning.

All models were pre-trained on existing datasets and then fine-tuned for this specific task. The final layer of each model was adjusted to classify the seven eye conditions.

2.3 Training Strategy

During training, images were passed through the models in batches. The models made predictions, and the difference between predicted and actual labels was calculated using a loss function. The Adam optimizer was used to update the model parameters because it helps the model learn efficiently. A small learning rate was chosen to ensure stable training. The models were trained for a limited number of iterations, and the best performing version was selected based on validation accuracy.

2.4 Evaluation Metrics

To measure performance, several evaluation metrics were used.

- Accuracy shows how many predictions were correct overall.
- Precision measures how many predicted cases are actually correct.
- Recall shows how well the model identifies all relevant cases.
- F1 score provides a balance between precision and recall.
- A confusion matrix was also used to analyze errors between different classes.

These metrics are important in medical applications because both false positives and false negatives can have serious effects.

2.5 Model Interpretability using Grad-CAM

To understand how the model makes decisions, a method called Grad CAM was used. This technique creates heatmaps that highlight important areas in the image that influence the prediction. These heatmaps were placed over the original retinal images to check whether the model focuses on meaningful regions such as lesions or abnormal structures. This helps ensure that the model is making reliable and meaningful predictions.

2.6 Experimental Setup

All experiments were carried out using a system with GPU support to speed up training. The models were implemented using the PyTorch framework. A batch size of 32 was used, and the data was shuffled during training to avoid bias. The overall process includes preparing the data, training the models, evaluating their performance, and analyzing the results. This approach allows a fair comparison between different models and helps identify the most effective one.

3. RESULTS AND DISCUSSION

3.1 Quantitative Performance Evaluation

The performance of the proposed deep learning framework was evaluated using four state-of-the-art convolutional neural network architectures: MobileNetV2, ResNet18, EfficientNet-B0, and DenseNet121. The models were assessed using classification accuracy as the primary metric. The experimental results show that DenseNet121 achieved the highest classification accuracy of 89.55%, followed by ResNet18 with 87.22%, EfficientNet-B0 with 86.37%, and MobileNetV2 with 84.69%. These results indicate that deeper architectures with improved feature propagation mechanisms perform better in capturing complex retinal patterns. DenseNet121 outperforms other models due to its dense connectivity structure, which enables effective reuse of features across layers and improves gradient flow. In contrast, MobileNetV2, while computationally efficient, demonstrates relatively lower performance due to its lightweight design.

Table 1: Comparison of classification accuracy for MobileNetV2, ResNet18, EfficientNet-B0, and DenseNet121 on the retinal image dataset.

S. No.	Model	Accuracy
1	Mobile Net	84.69
2	ResNet	87.22
3	EfficientNet	86.37
4	DenseNet	89.55

3.2 Class-wise Performance Analysis using Confusion Matrices

Confusion matrices were used to understand how well each model classified the seven eye diseases. They show which predictions were correct and where mistakes happened. MobileNetV2 performed well for common classes like diabetic retinopathy and normal cases. However, it made more mistakes between similar diseases, such as confusing diabetic retinopathy with normal cases and glaucoma with myopia. ResNet18 showed better performance than MobileNetV2, with more correct predictions overall. Still, some confusion remained between similar classes. Efficient Net B0 gave more balanced results across all categories. It handled both simple and complex patterns well, but a few errors were still present. DenseNet121 performed the best among all models. Most predictions were correct, and there were fewer mistakes. This shows that it is better at distinguishing between different eye diseases.

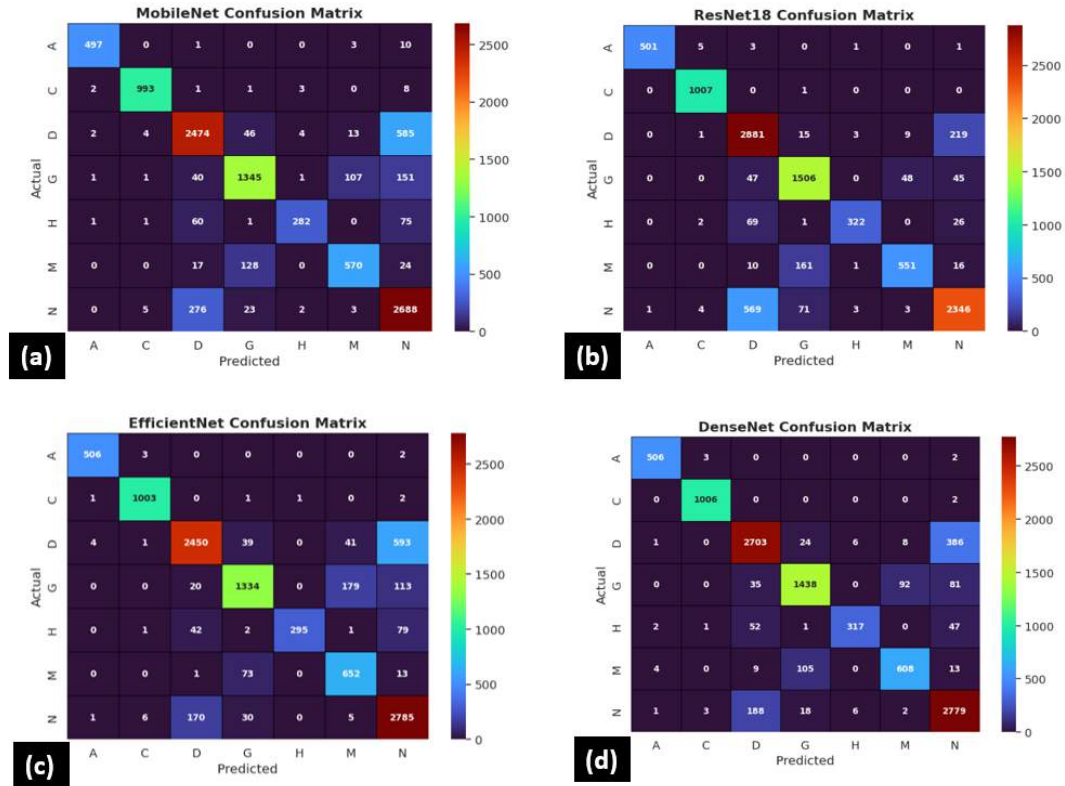


Figure 1: Confusion matrices of (a) MobileNetV2, (b) ResNet18, (c) EfficientNet-B0, and (d) DenseNet121, showing class-wise prediction performance across seven eye disease categories.

3.3 Model Interpretability using Grad-CAM

To improve transparency and interpretability, Grad-CAM was applied to the best-performing model (DenseNet121). This technique generates heatmaps that highlight the regions of the retinal image contributing most to the classification decision. The Grad-CAM visualizations reveal that the model focuses on clinically relevant regions such as lesions, abnormal blood vessels, and structural variations in the retina. For example, in diabetic retinopathy cases, the model emphasizes regions containing hemorrhages and exudates, while in glaucoma cases, attention is directed toward optic disc abnormalities. These results confirm that the model is not relying on irrelevant background features but is instead learning meaningful patterns associated with specific eye diseases.

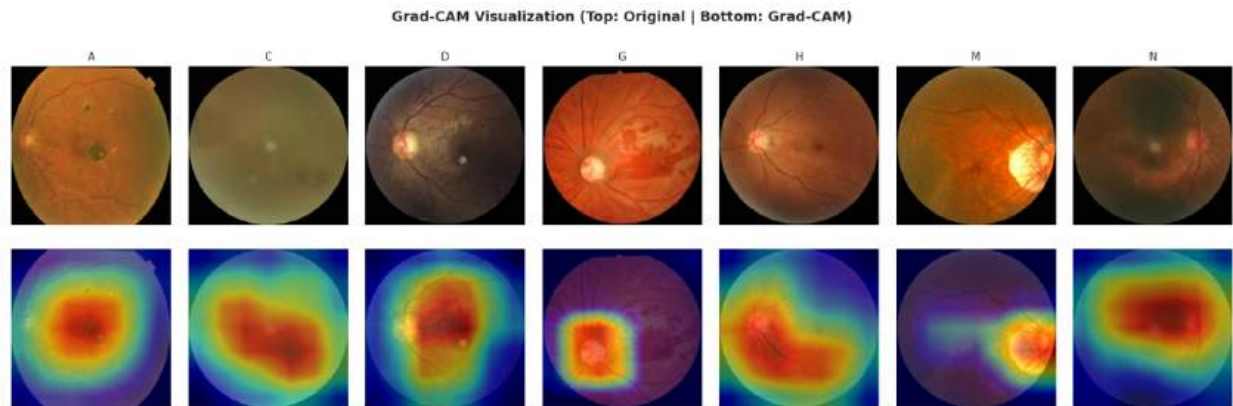


Figure 2: Grad-CAM visualization for seven eye disease classes. The top row shows original retinal images, while the bottom row shows corresponding heatmaps highlighting regions influencing model predictions.

The experimental findings demonstrate that model architecture plays a critical role in eye disease classification. DenseNet121 consistently outperforms other models due to its ability to efficiently reuse features and maintain strong gradient flow throughout the network. The comparison also highlights a trade-off between computational efficiency and accuracy. While MobileNetV2 is suitable for resource-constrained environments, it sacrifices some performance. On the other hand, DenseNet121 provides superior accuracy at the cost of increased computational complexity. The confusion matrix analysis indicates that most misclassifications occur between visually similar disease categories, which is a known challenge in retinal image analysis. However, the overall performance remains strong, with high accuracy and balanced classification metrics. The integration of Grad-CAM further strengthens the reliability of the system by providing visual evidence of model decision-making. This is particularly important in medical applications, where interpretability is essential for clinical acceptance. Overall, the proposed framework achieves strong performance in multi-class eye disease classification. The combination of high accuracy, detailed class-wise analysis, and interpretability demonstrates the effectiveness of deep learning models, particularly DenseNet121, in supporting automated diagnosis and large-scale screening systems.

4. CONCLUSION

In this study, a deep learning-based system was developed to classify multiple eye diseases using retinal images. Seven different conditions were considered, making the problem more realistic compared to simpler classification tasks. Four models were tested, and the results showed that all of them were able to learn useful patterns from the images. Among them, DenseNet121 achieved the highest accuracy of 89.55 percent and performed better than the other models. The results also showed that some errors still occur between diseases that look similar, which is a common challenge in this type of task. Grad CAM was used to understand how the model makes decisions. The heatmaps showed that the model focuses on important areas in the retina, such as lesions and structural changes. This helps increase confidence in the model's predictions. There is also a trade-off between accuracy and efficiency. Lightweight models like

MobileNetV2 are faster but less accurate, while deeper models like DenseNet121 provide better performance but require more computational power. This study has some limitations. It was tested on a single dataset, and the models were trained for a limited time. Future work could include using larger datasets, combining multiple models, and testing the system in real clinical settings. Overall, the results show that deep learning can be used effectively for detecting multiple eye diseases. This approach could support doctors and help enable faster and more accessible screening in real world applications.

Works Cited

- Baptista, Antonio M. G., Joel Monteiro, Marco Vieira, et al. 2015. "Causes of Vision Impairment in Portugal: A Hospital Based Study." *Investigative Ophthalmology & Visual Science*.
- Marques, Ana Patricia, Jacqueline Ramke, John Cairns, et al. 2021. "Global Economic Productivity Losses from Vision Impairment and Blindness." *EClinicalMedicine* 35. <https://doi.org/10.1016/j.eclinm.2021.100852>.
- Rajan, Karuna, and C. Sreejith. 2019. "Retinal Image Processing and Classification Using Convolutional Neural Networks." In *Lecture Notes in Computational Vision and Biomechanics*, vol. 30. https://doi.org/10.1007/978-3-030-00665-5_120.
- Swenor, Bonnielin K., and Joshua R. Ehrlich. 2021. "Ageing and Vision Loss: Looking to the Future." In *The Lancet Global Health*, vol. 9. no. 4. Preprint. [https://doi.org/10.1016/S2214-109X\(21\)00031-0](https://doi.org/10.1016/S2214-109X(21)00031-0).

**Not All Motor Skills Consolidate Equally:
Auditory-Motor Coupling and Sleep-Dependent Consolidation in Music Learning
By Elizabeth Zhao**

Sleep serves as an active neurobiological window for the consolidation of skills and memories acquired during waking hours. Decades of research have demonstrated that motor skills improve overnight without additional practice, a phenomenon that depends on specific neural events during non-rapid eye movement (NREM) sleep (Walker et al., 2002). Yet, much of this literature remains centered on a single type of task: sequential finger-tapping, in which participants press keys in a memorized order in isolation from sound, meaning, or aesthetic expectation. This has led to a tendency to treat motor skill consolidation as a relatively uniform process, largely independent of task context—that the neural processes stabilizing a pianist's newly learned passage are functionally identical to the rote acquisition of non-isochronous motor sequences, such as typing. This reductionist view overlooks the obligatory auditory-motor coupling that defines musical performance, where the motor trace is inextricably bound to an acoustic target.

However, this assumption is incomplete. Music is a motor skill, but it is fundamentally multimodal. When a musician practises a new passage, every keystroke is co-encoded with its anticipated auditory outcome, engaging the motor and auditory cortices within a real-time prediction-feedback loop (Zatorre, Chen & Penhune, 2007; Zatorre & Salimpoor, 2013). Early-stage musical practice and consciously learned sequences in music also likely engage explicit encoding processes more robustly than standard laboratory tasks, thereby increasing the likelihood of hippocampal-cortical recruitment.

Furthermore, music activates the brain's reward system in a way that typing tasks cannot: reward-related dopaminergic engagement during practice may increase the salience of the encoded memory trace, which in turn biases its likelihood of being selected for replay during subsequent sleep, thereby prioritising it for consolidation (Salimpoor et al., 2011; Ferreri et al., 2019). Consequently, the memory trace entering sleep is structurally distinct from a pure motor memory, representing an integrated auditory-motor-reward unit formed during encoding. This complexity poses a critical question that current sleep-memory literature has yet to fully resolve: when the sleeping brain replays musical memories, does it reactivate the full auditory-motor unit, or merely the mechanical motor component? Resolving this is essential not only for music cognition but for understanding the consolidation of any complex, multi-system sensorimotor memory. Music thus offers a uniquely tractable model for such inquiry, as the auditory and motor systems are deeply intertwined, and the behavioral consequences of their overnight consolidation can be quantified with high resolution. To address this gap, the present review first delineates the general neural mechanisms of motor consolidation, followed by an examination of how the auditory and reward systems modulate these processes. Finally, we propose a multi-system replay framework that accounts for the unique structural and functional properties of musical memory.

Sleep and Procedural Memory

Before examining what makes music special, it is necessary to understand what sleep does to motor memories in general.

Sleep Produces Overnight Motor Gains

The foundational evidence comes from Walker and colleagues (Walker et al., 2002), who demonstrated that a night of sleep produces a 20% increase in motor speed — a gain entirely absent after equivalent wakefulness — correlating positively with stage 2 NREM sleep duration. Walker (Walker et al., 2005) further showed that overnight gains do not depend on how much improvement was achieved during training, establishing sleep and practice as separable stages of memory formation. The irreversibility of the consolidation window was demonstrated by Fischer and colleagues (Fischer et al., 2002): participants sleep-deprived on the night following training failed to consolidate even after two subsequent recovery nights. Stickgold (Stickgold, 2000) confirmed that this sleep dependence extends beyond motor learning to perceptual skills, and that wakefulness — however prolonged — cannot substitute for it.

Enhancement or Stabilization? A Conceptual Debate

While these overnight gains are well-replicated, their interpretation remains contested. The dominant model frames sleep as an enhancement mechanism, producing performance levels that exceed anything achieved during training. Genzel and colleagues (Genzel et al., 2015) challenged this directly, showing that sleep primarily stabilizes an early performance boost that develops rapidly after training but decays across wakefulness — protecting existing gains rather than generating new ones. The distinction matters for musical memory in particular: it reframes sleep's role not as a mechanism for playing better the next morning, but as protection against the interference Allen (Allen et al., 2007) documented, stabilizing a newly learned melody against competing auditory-motor traces acquired later in the same practice session. It is worth noting that this enhancement-versus-stabilization debate has not been fully resolved. Some studies continue to report genuine overnight enhancement under conditions where post-training decay is controlled for, suggesting that the two functions — stabilization and enhancement — may not be mutually exclusive but may instead reflect different phases or mechanisms of the same consolidation process. For the purposes of this review, the distinction matters less than what both accounts agree on: sleep is uniquely necessary for protecting and extending motor learning in ways that wakefulness cannot replicate, and the window immediately following practice is the critical period in which this process is initiated.

Task Complexity Scales the Sleep Benefit

A further dimension concerns how strongly the benefit depends on task demands. Kuriyama (Kuriyama et al., 2004) showed that overnight improvement scales with complexity: simple unimanual sequences yielded approximately 17–18% gains, while a nine-element bimanual sequence produced a markedly greater 28.9% — more than twice the gain for the most

complex condition. This suggests that complexity is not a barrier to sleep-dependent consolidation but a catalyst: the more cortical systems engaged during encoding, the more sleep stands to offer during offline reprocessing. For music, which simultaneously recruits motor planning, auditory prediction, error monitoring, and rhythmic coordination, this scaling relationship suggests that sleep-dependent consolidation may be especially consequential.

Figure 1.

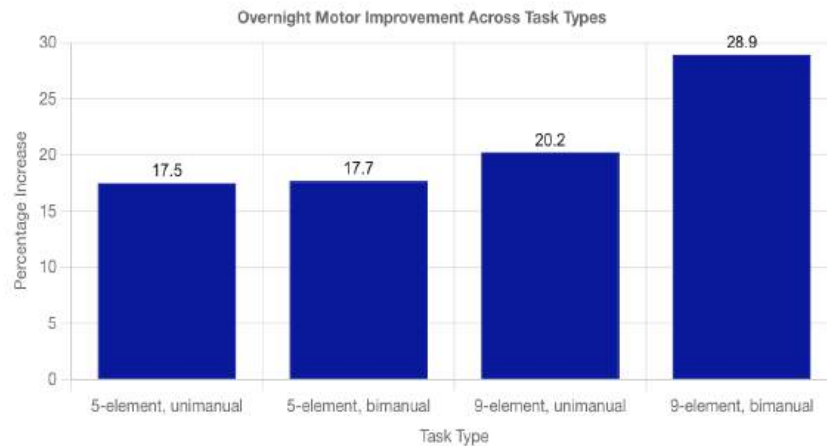
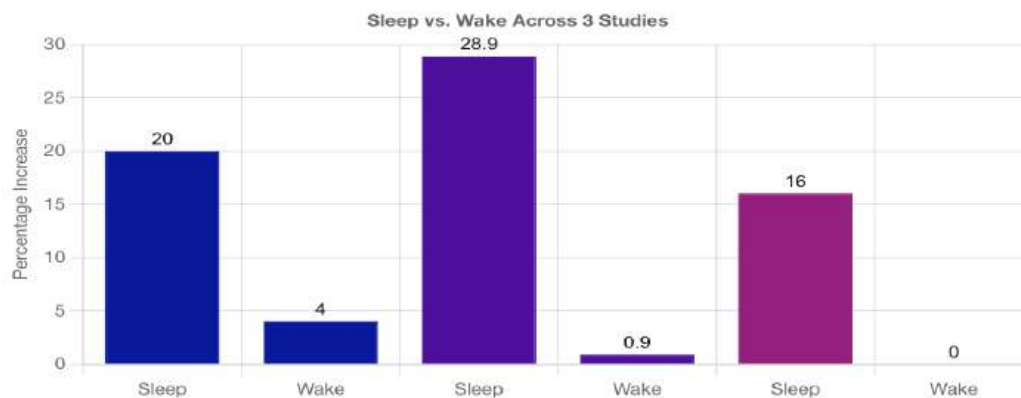


Figure 1. Overnight performance gains as a function of task complexity. Adapted from Kuriyama et al. (2004).

As illustrated in Figure 1, the disproportionate overnight gain in complex bimanual tasks (28.9%) suggests that as motor demands increase and cortical recruitment broadens, the brain becomes increasingly dependent on sleep to integrate these distributed neural networks. The asymptotic performance level reached post-sleep further suggests that complexity is not a barrier to sleep-dependent consolidation but a catalyst for it — the more systems engaged during encoding, the more sleep stands to offer during offline reprocessing.

Figure 2



Walker et al. (2002) — simple tapping Kuriyama et al. (2004) — complex bimanual Nishida & Walker (2007)
— nap paradigm

Figure 2. Illustrative comparison of overnight performance gains reported across three motor-learning studies. Because these studies used different tasks and designs, the figure is intended as a conceptual summary rather than a direct quantitative comparison.

Taken together, the evidence reviewed in this section establishes three converging principles: sleep is uniquely necessary for motor memory consolidation in a way that wakefulness cannot replicate; its primary function may be stabilization of early gains rather than linear enhancement; and the magnitude of the benefit scales with the complexity and cortical breadth of the encoded task. These principles set the foundation for the central argument of this review — that musical motor learning, by virtue of its multimodal structure, explicit demands, and reward engagement, may engage sleep-dependent consolidation more deeply and across a broader network of cortical systems than any standard laboratory motor task.

The Neural Machinery: Spindles, Slow Oscillations, and Hippocampal Replay

The behavioral gains described above are produced by specific neural events during NREM sleep. Diekelmann and Born (Diekelmann et al., 2010) provided the most influential theoretical account: during slow-wave sleep (SWS), cortical slow oscillations (SOs) coordinate with thalamocortical sleep spindles to orchestrate the consolidation of newly encoded memories. Born and Wilhelm (Born et al., 2012) elaborated this model, demonstrating that SWS supports hippocampal-cortical memory transfer, while subsequent REM sleep contributes to integration and stabilization.

The role of the hippocampus in motor and procedural memory consolidation is more nuanced than in declarative memory. Sleep-dependent consolidation of motor sequence learning may recruit hippocampal-cortical dialogue under some conditions – particularly when tasks are explicitly learned, sequence-based, or involve the kind of episodic encoding associated with attentional engagement (Albouy et al., 2008; Rasch & Born, 2013). For purely implicit, automatic motor skills, hippocampal involvement is less certain. This distinction is critical for music, where learning encompasses a spectrum from highly explicit, consciously rehearsed passages to increasingly implicit and automatized movements developed through extended practice.

Boutin and Doyon (Boutin et al., 2020) refined the role of spindles within this framework. They proposed that what matters is not merely spindle density but the organization of spindles into temporal clusters called “trains” – groups of consecutive spindles separated by precisely timed refractory periods oscillating at approximately 0.2-0.3 Hz. These spindle trains provide a structured temporal window for reprocessing motor memory traces. Fogel and Smith (Fogel et al., 2011) further reviewed evidence that spindle density may serve as a physiological index of learning capacity: individuals who produce more frequent and organised spindles show greater overnight improvement and perform better on cognitive assessments. This raises the possibility

that years of musical training, which extensively reshape neural architecture, may also refine these spindle characteristics to support more efficient consolidation.

The most comprehensive recent synthesis comes from Klinzing, Niethard and Born (Niethard et al., 2019), who describe the full SO-spindle-ripple coupling hierarchy: slow oscillation up-states create temporal windows for spindle generation; spindles temporally nest hippocampal sharp-wave ripples (SWRs); and SWRs carry temporally compressed neural replay of waking experience to cortical networks for long-term storage. Joo and Frank (Joo et al., 2018) reviewed the evidence that SWRs are the neural substrate of this replay, showing that selectively disrupting them during sleep impaired subsequent memory performance – establishing a causal role for hippocampal replay in long-term memory formation. Importantly, while this hierarchy has been primarily studied in the context of individual motor or declarative traces, musical learning challenges this unimodal framework by requiring the simultaneous reactivation of integrated auditory and motor networks.

Whether the SO-spindle-ripple cascade is capable of coordinating such multi-system replay — and whether spindle activity localized over both motor and auditory cortex is a necessary condition for complete musical memory consolidation — are precisely the questions that the remainder of this review works toward. The machinery is well characterized; what remains unresolved is whether this well-characterized machinery can coordinate the simultaneous reactivation of multiple interdependent systems, as required for musical memory.

What Makes Musical Motor Learning Different

Herholz & Zatorre (Herholz et al., 2012) and Schlaug (Schlaug, 2015) both explore how musical motor learning differs from the tasks mentioned above in three fundamental respects. First, it is multimodal rather than purely motor: every action is bound to an auditory representation from the first moment of practice. Second, it is predictive rather than merely reactive: the auditory system generates forward models of expected sound that constrain and guide motor output in real time. Third, in its early explicit phase, musical practice is rewarding and consciously attended in ways that may bias the encoding and subsequent overnight consolidation of musical memories. These are not simply differences of degrees; they are differences in the kind of memory that is formed.

Music Learning Is Multimodal, Not Merely Motor

The foundational account of auditory-motor coupling in music comes from Zatorre, Chen, and Penhune (Zatorre et al., 2007), who proposed that music performance is mediated by a continuous bidirectional loop between motor output and auditory feedback, running through the dorsal auditory stream, posterior parietal cortex, and premotor cortex. During performance, these pathways operate simultaneously: every intended movement generates an auditory prediction before execution, and auditory feedback from the resulting sound updates the motor plan for the next movement. Music is therefore not a motor task with auditory consequences added afterward; it is a unified sensorimotor action–perception cycle.

Pranjć and colleagues (Pranjć et al., 2024) confirmed and expanded this picture through a systematic scoping review of neuroimaging studies of auditory-motor synchronization. Their synthesis found converging evidence that the neural network supporting sound-to-movement coupling is broader and more bilateral than earlier models suggested, engaging the supplementary motor area, premotor cortex, ventrolateral prefrontal cortex, basal ganglia, and cerebellum simultaneously. Beta-band oscillations (13–30 Hz) play a key role in linking auditory cues to motor output in real time, and the degree of this coupling is modulated by rhythmic complexity.

The depth of this coupling in trained musicians is further demonstrated by Haueisen and Knösche (Haueisen et al., 2001), who showed that simply listening to known music activates the motor cortex in trained pianists even without any movement. Magnetoencephalographic recordings revealed involuntary motor activity — the motor system generating representations of the keystrokes associated with the heard music, without any intention to move. This finding illustrates that auditory and motor representations have become structurally fused through years of co-activation: the boundary between hearing and doing is neurologically blurred, and a musical memory is therefore not a motor memory with an audio track attached but a unified multimodal trace.

This structural coupling is reflected in the physical anatomy of the musician's brain. Schlaug (Schlaug, 2015) reviewed evidence that repeated auditory-motor associations strengthen connections between auditory and motor cortex, particularly the arcuate fasciculus. Wan and Schlaug (Wan et al., 2010) provided direct neuroimaging evidence: musicians show significantly greater arcuate fasciculus volume and higher fibre coherence compared to non-musicians, correlating with the duration and intensity of prior training — structural rewiring rather than transient functional coupling.

Behavioral evidence from Palmer and Mathews (Palmer et al., 2013) and Bernardi and colleagues (Bernardi et al., 2013) further supports this view: even in the absence of physical execution, auditory imagery specifically predicts the degree of movement anticipation, confirming that the auditory and motor systems remain co-engaged during musical learning and reinforcing the idea of a unified multimodal representation.

Music Learning Is Explicit, Rewarding, and Overlearned: The Role of the Reward System

A second set of dimensions concerns the motivational and attentional conditions of musical practice. Early-stage practice engages explicit encoding more strongly than standard laboratory tasks, recruiting the hippocampal-cortical system that mediates sleep-dependent consolidation (Robertson et al., 2004). Unlike a finger-tapping sequence, which can be executed with diminishing conscious involvement after only minutes of practice, a new musical passage demands sustained error monitoring across hundreds of repetitions because each movement must match both a motor target and an auditory expectation simultaneously. As skills become automatised through extended practice they take on increasingly implicit and habitual characteristics, so the relationship between explicit encoding and sleep benefit likely varies

across the course of musical skill acquisition — a nuance that future studies comparing early and late stages of musical learning would need to address.

Beyond explicitly attentional engagement, music activates the brain's reward system in a way that a typing task categorically does not. When music meets expectation, such as when a harmonic resolution arrives after tension, or a familiar melodic shape is recognized, the brain releases dopamine in the striatum, producing the pleasurable emotional responses associated with musical experience. Salimpoor and colleagues (Salimpoor et al., 2011) used PET neuroimaging with a dopamine-specific radioligand to demonstrate anatomically distinct dopamine release in both the caudate (during anticipation of a pleasurable musical moment) and the nucleus accumbens (during the peak emotional response itself). The causal role of dopamine in this process was suggested by Ferreri and colleagues (Ferreri et al., 2019), who showed that pharmacological manipulation of dopamine availability directly modulates the degree of musical pleasure experienced.

Importantly, this reward–prediction coupling appears to be a consistent feature of musical experience rather than an incidental byproduct. Gold and colleagues (Gold et al., 2023) found that the interaction between musical uncertainty and surprise—the moment of prediction confirmation or violation—engages both auditory and reward structures (STG and ventral striatum). As Rasch and Born (Rasch et al., 2013) synthesized, reward-related encoding biases the prioritization of memories for overnight consolidation. The brain preferentially reprocesses those traces tagged as motivationally significant. Consequently, every prediction-confirmation cycle during music practice generates a dopaminergic signal that may "tag" the musical memory for later consolidation. However, direct evidence linking dopaminergic reward signals during musical practice to subsequent sleep-dependent replay remains limited. This remains a compelling hypothesis: that dopamine tagging during musical practice increases the likelihood of overnight replay, but one that awaits direct empirical confirmation.

Music Learning is Predictive, Not Merely Reactive

A third dimension that distinguishes musical motor learning is the role of prediction. In typical finger-tapping experiments, participants receive reactive feedback. In music, however, the auditory cortex operates prospectively: it generates forward predictions about upcoming sounds based on learned musical structure, monitoring whether the actual outcome matches the expectation (Zatorre & Salimpoor, 2013).

This predictive structure has direct implications for memory encoding. Over many repetitions, what is stored is not merely a sequence of movements but a set of auditory-motor prediction-confirmation loops — each movement linked to its expected sound, and each sound linked back to the movement that produced it. The resulting representation is bidirectional: activating the motor trace should activate the corresponding auditory prediction, and hearing the sound should activate the corresponding motor plan. This bidirectional structure is what makes musical memories accessible to auditory cues during sleep.

Taken together, reward-related dopaminergic engagement during music practice may increase the salience and prioritization of musical memories for later consolidation. This remains a hypothesis — the prediction that dopamine tagging during musical practice increases the likelihood of overnight replay has not yet been directly tested.

Evidence: What Happens When Musicians Sleep

Musical memories exhibit sleep-dependent consolidation at least as robustly as generic motor tasks, yet they display distinctive features—such as accuracy-specific gains and vulnerability to auditory interference—that reflect their integrated auditory-motor structure.

Sleep Enhances Musical Performance

Allen (Allen, 2007) trained sixty experienced non-pianist musicians to perform a 12-note piano melody during evening practice sessions, finding significant overnight gains in correct key presses without any additional practice — reflecting genuine consolidation rather than continued learning. This suggests that sleep-dependent consolidation extends to real musical material, though ecological validity remains limited.

Simmons and Duke (Simmons et al., 2012) identified a critical dissociation between accuracy and speed: while speed can improve over short waking intervals, significant accuracy gains emerge only after a 24-hour period including sleep. This finding is theoretically pivotal: in music, accuracy is the dimension most tightly coupled to auditory feedback monitoring. That accuracy—the very heart of sensorimotor matching—is specifically sleep-dependent suggests that the "auditory-monitoring" component of the memory trace is the primary beneficiary of nocturnal processing.

Lewis and Durrant (Lewis et al., 2011) broadened the evidence base by testing overnight consolidation of rhythmic motor learning. Their neuroimaging results revealed that striatum and dorsal cerebellum responses correlated specifically with overnight improvement in rhythmic performance after sleep but not after an equivalent waking interval, with additional post-sleep hippocampal activation observed in auditory monitoring conditions. Rhythm is the temporal backbone of musical motor performance, and this study shows that it too requires sleep and engages a subcortical network.

Fischer and colleagues (Fischer et al., 2002) established that sleep does not merely enhance motor memories but forms them: when participants were sleep-deprived immediately after training, they failed to consolidate even after two nights of subsequent rebound sleep. Applied to music, this implies that the consolidation window for a newly learned passage is narrow and irreversible.

Across studies, a consistent pattern emerges: musical consolidation is not only robust, but selectively enhances aspects of performance most tightly linked to auditory–motor integration. However, it is important to note that most of these existing studies rely on relatively simplified musical tasks, often involving short sequences learned over a single session. This raises concerns about ecological validity, as real musical practice typically involves extended, multi-day

learning, expressive variation, and bimanual coordination. As such, the extent to which these findings generalize to real-world musical expertise remains uncertain.

Sleep Stabilizes Musical Memories Against Interference

A critical conceptual distinction for this section is between enhancement and stabilization. Genzel and colleagues (Genzel et al., 2015) challenged the straightforward enhancement model, providing evidence that sleep primarily stabilizes performance — protecting the rapid initial gains from post-training decay and interference — rather than generating additional enhancement beyond any level seen before sleep. Under this framework, the most important function of sleep for a musician may be less about playing faster the next morning and more about ensuring that what was learned is not overwritten by subsequent practice.

The most revealing evidence for the auditory-motor structure of musical memory comes from interference studies. Allen (Allen, 2007) included a condition in which musicians learned a second melody immediately after the target melody before sleeping. Participants in this condition showed no overnight gains whatsoever on the target melody — consolidation was completely abolished. This mirrors the classical motor interference finding of Brashers-Krug, Shadmehr and Bizzi (1996), which showed that learning a second motor task shortly after the first blocks consolidation within a critical window.

The key question is the nature of this interference. In Brashers-Krug et al., the two sequences were different motor patterns with no shared auditory features. In Allen's music study, the two melodies differed in both motor patterns and auditory profiles. One possibility is that the interference is entirely motor; another is that the shared auditory format of the two melodies prevented the brain from tagging them as distinct consolidated traces, because both activated overlapping auditory prediction networks during encoding. Interestingly, Allen (Allen, 2013) showed a reactivation-rescue effect: a brief pre-sleep re-practice of the first melody can restabilize the trace, effectively shielding it from the interference of the second melody during subsequent sleep.

A fundamental ambiguity, however, remains unresolved. No study has held motor overlap constant while varying auditory overlap, or vice versa — for example, by pairing melodies with identical pitch content but different motor patterns, or identical motor patterns mapped onto different pitch sequences. This dissociation is not merely a methodological footnote: it is the key experiment that would most directly test whether musical memories are stored as unified auditory-motor traces or as parallel but separable representations. Until it is conducted, the interference pattern remains consistent with the auditory-motor integration hypothesis but cannot be taken as confirmation of it.

Musical Expertise Reshapes the Consolidation System Itself

If musical training physically rewires the auditory-motor system, one would expect musicians to show a consolidation advantage not only on musical tasks but on generic motor

tasks that share structural features with musical performance. Tucker and colleagues (Tucker et al., 2016) tested this directly, comparing musicians and non-musicians on overnight consolidation of a sequential typing task with no musical content. Musicians showed 23.7% improvement over 12 hours compared to only 11.0% for non-musicians, and this advantage held even when controlling for superior initial performance at training. More strikingly, musicians retested after a day of wakefulness performed slightly better than non-musicians who had slept, suggesting that the musically trained brain can consolidate motor skills across wakefulness in a way that non-musicians cannot achieve even with a full night of sleep.

The musician advantage is not simply a function of better initial encoding; it appears to reflect a genuine difference in the offline consolidation machinery itself, though whether this stems from structural white-matter differences, altered spindle organisation, or expanded cortical circuits for explicit motor learning remains unclear. The implication is significant: musical training may shape not only what is learned but how efficiently it is consolidated during sleep. Nishida and Walker (Nishida et al., 2007) provided evidence that motor memory consolidation during even a brief nap is possible and regionally specific. Participants who napped after a motor learning session showed 16% consolidation enhancement relative to those who remained awake, and the magnitude of improvement correlated with spindle activity in the learning hemisphere. The spindles that drive motor consolidation map onto the cortical representation of the learned skill. If musical training shifts or expands this cortical representation, it may also shift or expand the topography of the spindle activity that drives consolidation.

What Neural Signatures Do We Expect

While the neural mechanisms of musical consolidation have not yet been directly imaged, current frameworks allow for specific predictions. Barakat and colleagues (Barakat et al., 2011) established that fast sleep spindles (13-15 Hz) over motor cortex predict overnight gains on generic motor sequences, and that these spindles increase in density following motor learning. Robertson and colleagues (Robertson et al., 2018) showed, using TMS disruption, that overnight consolidation of explicitly learned motor sequences specifically requires intact primary motor cortex (M1), while implicitly learned sequences require the inferior parietal lobule (IPL). Because musical practice in its early phase engages explicit encoding processes, musical motor memories should route through the M1-dependent consolidation circuit and be reflected in spindle activity over the motor cortex.

However, if musical memories are stored as auditory-motor coupled representations, then the spindle-mediated reprocessing of those memories during sleep should also engage the auditory cortex. Boutin and colleagues (Boutin et al., 2023) showed that spindle trains underlie motor consolidation regardless of whether the skill was learned through physical practice, motor imagery, or observation, suggesting that the mechanism is robust across learning modalities. Since musicians inherently combine these modes, we predict that nocturnal spindles will exhibit a distributed topography across both motor and auditory cortices, reflecting the multi-system structure of the original encoding process.

Does Musical Memory Replay Involve the Auditory Cortex

The reviewed evidence demonstrates that musical motor memories consolidate during sleep and that this process has characteristics consistent with the multimodal structure of musical encoding. Yet, the central question – whether the auditory cortex participates alongside the motor cortex in the overnight replay of musical memories – has not yet been directly answered. While the current evidence does not directly demonstrate such participation, several lines of evidence make it a plausible hypothesis.

Musical memories are encoded as auditory-motor integrated traces

The case that musical memories are stored as auditory-motor integrated units, rather than as parallel but independent auditory and motor traces, is supported by the structural and functional evidence referenced previously. Haueisen and Knösche (Haueisen et al., 2001) showed that hearing known music activates motor representations in pianists; Zatorre and Salimpoor (Zatorre et al., 2013) showed that motor output generates auditory predictions; Wan and Schlaug (Wan et al., 2010) showed that the white-matter tract connecting auditory and motor cortex is enlarged in musicians. Bernardi and colleagues (Bernardi et al., 2013) demonstrated that mental practice in musicians engages auditory imagery — specifically the anticipation of upcoming sounds — and that this auditory component independently predicts the degree of motor anticipation that develops. These findings are consistent with an encoding architecture in which auditory and motor information are stored in a tightly linked representation. These findings are consistent with an encoding structure in which auditory and motor information are stored in a shared or tightly linked representation rather than in separate, independently consolidable stores.

If this is correct, then disrupting one component of the memory should affect the other. The interference evidence in Allen (Allen, 2007) is at least consistent with this prediction: learning a second melody disrupted consolidation of the first in a way that was more complete than what is typically observed with purely motor interference. The dissociation between accuracy and speed in Simmons and Duke (Simmons et al., 2012) — accuracy being specifically sleep-dependent, speed less so — may reflect the differential dependence of the auditory and motor components on overnight consolidation. None of this is definitive, but the pattern is coherent with an integrated encoding model.

TMR studies suggest that musical memories can be accessed through auditory cues

The most direct human evidence bearing on the auditory-motor integration hypothesis comes from targeted memory reactivation (TMR). The approach relies on presenting sensory cues associated with a prior learning experience during sleep, to determine whether those cues can selectively bias the consolidation of the associated memory. Rudoy and colleagues (Rudoy et al., 2009) established that sounds associated with specific objects during spatial learning, when played quietly during SWS, improved memory for the cued objects. The mechanism is thought to be that the auditory cue reaches the sleeping brain and biases hippocampal SWR replay toward

memories associated with that cue – as Bendor and Wilson (Bendor et al., 2012) directly demonstrated in rats, where auditory cues during sleep selectively increased hippocampal replay of memories in the presence of those sounds.

A critical finding in this context comes from the contrast between Rasch and colleagues (Rasch et al., 2007) and Antony and colleagues (Antony et al., 2012). Rasch et al. showed that an odor cue presented during SWS improved retention of a hippocampus-dependent spatial memory but had no effect on performance on a generic finger-tapping procedural task. The sleeping brain could not be cued to reactivate a purely motor memory through a contextual sensory cue. Antony et al., by contrast, showed that melody tones played during SWS selectively improved performance on the cued melody the following morning, both in terms of accuracy and in terms of EEG signatures of memory processing during the cue presentation period.

Recent methodological advances have begun to characterize the neural properties of this reactivation more precisely. Rakowska and colleagues (Rakowska et al., 2025) directly measured the temporal properties of memory reactivation following TMR cues using EEG and machine learning classification, finding that reactivation occurred 3 to 20 times faster than waking activity and was more frequently detected in trials with high sleep spindle power — providing direct evidence for the positive relationship between spindle activity and the detectability of memory reactivation during sleep. This temporal compression is a hallmark of neural replay, and its presence during auditory cuing suggests that the auditory cortex may act as a gateway to the broader sensorimotor memory.

A critical interpretive limitation of the TMR paradigm, however, deserves explicit acknowledgement. The behavioral improvement produced by auditory cueing during sleep is consistent with auditory cortex involvement in the reactivation process, but it is equally consistent with an account in which the cue triggers reactivation entirely within motor circuits that happen to be associatively linked to the auditory cue — with the auditory cortex playing no active role during sleep itself. The behavioral outcome alone cannot distinguish between these two mechanistic accounts. What is needed is simultaneous measurement of both motor and auditory cortical activity during cue-triggered reactivation in sleeping musicians — a technically demanding but feasible experiment given current high-density EEG and source localization methods. Until such data exist, the TMR evidence is best understood as establishing that musical memories have an auditory retrieval handle accessible during sleep, without specifying where in the brain that handle is held.

Hu and colleagues (Hu et al., 2020) confirmed through meta-analysis of 91 TMR studies that TMR during NREM sleep reliably improves memory across declarative, procedural, and skill domains. The effect is statistically strong and has been replicated across many laboratories. Schönauer, Geisler and Gais (Geisler et al., 2014) showed that auditory TMR of motor sequences correlates with fast spindle activity over motor regions, suggesting that cue-driven reactivation works by triggering spindle-mediated consolidation in task-relevant cortical circuits. If auditory TMR of musical memories triggers spindle activity over both motor and auditory cortex, this would be direct support for integrated replay.

Taken together, these findings suggest that auditory cues may provide a privileged access point to musical memories during sleep, in contrast to purely motor representations.

Songbird Work Provides a Biologically Plausible Model of Integrated Replay

The most mechanistically concrete evidence for auditory participation in overnight replay comes from the songbird literature. Dave and Margoliash (Dave et al., 2000) recorded from neurons in the high vocal center (HVC) — a premotor structure critical for song production — in adult zebra finches during sleep, finding that HVC neurons spontaneously replayed learned song sequences with millisecond-level precision, without any external trigger. Elmaleh and colleagues (Elmaleh et al., 2021) extended this with high-density recordings, showing that replay spans syllable boundaries — the premotor circuit replays the learned song as a unified sequence, not a series of isolated motor commands.

The complementary finding concerns auditory responsiveness during the same sleep stage. A study from the Margoliash laboratory found that HVC auditory responses to playback of the bird's own song were strongest during slow-wave sleep — the precise stage in which motor replay occurs. The auditory system is not switched off during SWS; it is selectively active and maximally responsive. Taken together, these two findings establish that premotor replay and auditory responsiveness can co-occur during SWS in a learned vocal-motor skill, though whether they are causally linked or merely coincident remains an open question.

These findings require cautious translation to human musical memory, as the zebra finch song system is a specialized vocal-learning circuit that may not map directly onto keyboard motor learning. Nevertheless, as an existence proof of integrated sensorimotor replay during sleep, the songbird model provides the closest available biological analogue for what auditory-motor replay in sleeping musicians might look like.

A multisystem replay hypothesis for human musical memory

Recent advances in human neuroimaging during rest and sleep have begun to reveal that memory replay is not a purely hippocampal event. Higgins and colleagues (Higgins et al., 2021) used MEG and machine learning to detect sequential replay bursts in resting humans and found that these replay events were packaged into transient bursts occurring selectively during activation of the default mode network (DMN) and parietal alpha networks — networks believed to support inwardly oriented attention and to inhibit bottom-up sensory processing. The finding suggests that replay events are temporally orchestrated by whole-brain network states that create interference-free windows for internally generated memory reinstatement. Crucially, the replay-associated activity included temporal areas, consistent with the possibility that sensory cortical regions, including auditory cortex, may participate in replay of learned material that has sensory content. This provides a candidate neural mechanism for how the auditory cortex might be engaged during musical memory replay during sleep — through DMN-coordinated bursts that transiently suppress ongoing sensory input while simultaneously reinstating distributed cortical representations.

Human Musical Replay May Involve the Auditory Cortex

Drawing together the preceding lines of argument, a coherent picture emerges. Musical memories appear to be encoded as auditory-motor integrated traces, in which auditory and motor representations are not merely co-activated but structurally linked. These integrated traces seem to be accessible to auditory cues during sleep in a way that purely motor memories are not, suggesting that the auditory component of the memory provides a retrieval handle that the auditory component may provide a retrieval pathway through which the sleeping brain can access the integrated trace. In songbirds, premotor replay of learned vocal sequences during SWS co-occurs with auditory responsiveness in the same circuit. Together, these lines of evidence are consistent with the hypothesis that human musical memory replay during SWS involves the auditory cortex alongside the motor cortex.

This hypothesis is, however, currently inferential. The most direct test would involve high-density EEG in musicians following practice, measuring whether sleep spindles appear over auditory cortical regions (e.g., STG and planum temporale) in addition to M1, and whether the density of auditory spindles specifically predicts accuracy gains. Andrillon and colleagues (Andrillon et al., 2020) demonstrated that auditory stimulation can enhance delta-sigma coupling in spatially circumscribed regions. The primary limitation of current work is the lack of causal evidence linking auditory cortex reactivation directly to behavioral improvement in humans.

A Multisystem Replay Model of Musical Memory Consolidation

Core Hypothesis

This review proposes that musical memory should be conceptualized as an integrated auditory–motor–reward representation, rather than a purely motor trace. During sleep, we hypothesize that:

- (1) motor cortical regions replay action sequences,
- (2) auditory cortical regions simultaneously reinstate predicted sensory outcomes,
- (3) reward-related dopaminergic tagging biases which traces are prioritized for replay.

This multisystem replay process may be coordinated through spindle–slow oscillation coupling, resulting in preferential stabilization of accuracy-dependent aspects of musical performance.

Three Stage Walkthrough

The conceptual model below summarizes the framework proposed in this review. It distinguishes three stages: encoding, sleep-dependent consolidation, and behavioral outcome. Crucially, it treats musical memory as an integrated trace linking motor commands, auditory prediction and feedback, and reward salience.



Conclusion

This review has argued that music is not simply a harder version of finger-tapping. It is a qualitatively different kind of motor learning — one in which every movement is co-encoded with its anticipated auditory outcome, the encoding process is explicitly conscious and attention-demanding in its early phase, and the experience of practice is rewarding in a way that may tag the resulting memories for preferential overnight consolidation.

At the general level, sleep-dependent motor memory consolidation is well established. A single night of sleep produces motor gains that cannot be achieved through equivalent wakefulness; those gains are driven by NREM sleep spindles, slow oscillations, and hippocampal sharp-wave ripples operating in a precisely coordinated coupling hierarchy; and the gains are irreversible if sleep is missed in the critical window following learning. Complexity favors greater sleep benefit, and explicit encoding favors preferential access to the hippocampal-cortical consolidation process.

At the level of music specifically, the evidence shows that musical motor memories consolidate during sleep, that accuracy is specifically sleep-dependent in a way that speed is not, and that musical training reshapes the consolidation system itself such that musicians consolidate motor skills more efficiently than non-musicians even on tasks with no musical content. The interference pattern is consistent with the auditory-motor structure of musical memory. Taken together, these findings are consistent with the view that musical memories are formed and consolidated as integrated auditory-motor units, but this remains a hypothesis rather than an established conclusion.

The open question identified by this review is whether the auditory cortex participates in the overnight replay of musical memories. The songbird literature shows that premotor replay of learned vocal sequences occurs during SWS and co-occurs with auditory responsiveness in the same circuit. The human TMR literature shows that auditory cues during SWS can selectively reactivate musical motor memories, suggesting those memories have an auditory handle accessible during sleep. Still, direct neural evidence of auditory-cortical involvement in musical memory replay during human sleep does not yet exist. Addressing this question will require high-density EEG or related methods in trained musicians.

More broadly, music offers neuroscience a uniquely tractable model for studying multi-system sensorimotor consolidation — one where the auditory and motor components of learning are both precisely measurable, their coupling varies systematically with training, and the behavioral consequences of overnight consolidation can be quantified with high resolution. This framework suggests that sleep-dependent consolidation may operate fundamentally differently for memories that are inherently multisystem, challenging the traditional separation between procedural and declarative domains. Ultimately, understanding whether the sleeping brain replays the full auditory-motor experience of music, rather than isolated motor components, may reshape our broader understanding of how complex, multisystem memories are consolidated across distributed neural networks — and in doing so, reveal something fundamental about the offline machinery that transforms experience into lasting skill.

Limitations and Future Directions

Several limitations of the current evidence base deserve acknowledgement. First, most sleep and motor memory studies — including those reviewed here — use simplified laboratory tasks that share only surface features with real musical practice. Even Allen (Allen, 2007), the most ecologically relevant study, used short single-hand melodies learned over one evening; this is a far cry from the bimanual coordination, expressive shaping, and extended rehearsal that characterise professional musical practice. Whether the consolidation patterns observed with these simplified tasks scale to the complexity of real repertoire remains an open empirical question.

Second, the "musician advantage" literature, exemplified by Tucker and colleagues (Tucker et al., 2016), draws almost exclusively from keyboard players trained in the Western classical tradition. Different instruments make qualitatively different demands on the auditory-motor system: brass and woodwind players must coordinate breath pressure and embouchure with pitch anticipation; string players must continuously modulate bow pressure against an auditory target with no frets to anchor pitch. Whether these distinct sensorimotor profiles produce different patterns of sleep-dependent consolidation — and whether the auditory-motor integration hypothesis proposed here holds equally across instruments — is entirely unknown.

Third, the direct neural evidence for the central claim of this review is absent. No study has recorded EEG or neuroimaging data from musicians during natural sleep following

ecological musical practice and tested whether the auditory cortex participates in overnight replay. The behavioural and indirect evidence reviewed here is consistent with this hypothesis but cannot confirm it.

These limitations demonstrate that the most pressing experiment remains to be the high-density EEG in musicians following real musical practice, measuring whether sleep spindles appear over auditory cortex in addition to motor cortex, and whether auditory-region spindle density predicts overnight gains in pitch accuracy specifically. Beyond this, three further directions emerge from the framework developed here.

First, if musical memories are encoded as auditory-motor prediction-confirmation loops, sleep should selectively consolidate the corrections a musician makes during practice — the moments where a prediction error triggered a motor update — rather than distributing gains evenly across the passage. A fine-grained analysis of which specific transitions improve overnight, compared to which transitions generated the most prediction errors during evening practice, would provide a direct test of the predictive encoding hypothesis. This extends Kuriyama and colleagues' (Kuriyama et al., 2004) finding that sleep preferentially benefits the most difficult transitions, but grounds it in a mechanistic account of why difficulty predicts overnight gain.

Second, the present review has drawn its evidence almost exclusively from keyboard players in the Western classical tradition. Different instruments make qualitatively different demands on the auditory-motor system: wind players produce pitch entirely through breath and embouchure modulation against an auditory target, with no tactile anchor; string players continuously adjust bow pressure against a sound they are simultaneously predicting. If the degree of auditory-motor coupling during encoding drives the pattern of sleep-dependent consolidation, then instrument families should differ systematically in overnight gains and in the cortical topography of consolidation-related spindle activity. Comparing across instrument families would test whether the framework proposed here generalises beyond the keyboard paradigm that has dominated the literature.

Third, and perhaps most elegantly, the reward-tagging hypothesis can be tested without pharmacology by exploiting naturally occurring individual differences in musical reward sensitivity. Ferreri and colleagues (Ferreri et al., 2019) showed that dopamine modulates music-driven memory improvements, and a substantial minority of the population experiences music with little to no reward response despite normal perception — so-called musical anhedonia. If dopaminergic tagging during practice prioritises musical memories for overnight consolidation, then individuals with low musical reward sensitivity should show attenuated sleep-dependent gains on musical tasks relative to generic motor tasks, while showing no such attenuation on non-musical motor sequences. This would constitute the cleanest available test of whether reward engagement during encoding shapes the efficiency of subsequent sleep consolidation for music specifically.

Works Cited

- Allen, S. E. (2007). Procedural memory consolidation in musicians [Doctoral dissertation, University of Texas at Austin].
- Allen, S. E. (2013). Memory stabilization and enhancement following music practice. *Psychology of Music*, 41(6), 794–803.
- Antony, J. W., Gobel, E. W., O'Hare, J. K., Reber, P. J., & Paller, K. A. (2012). Cued memory reactivation during sleep influences skill learning. *Nature Neuroscience*, 15(8), 1114–1116. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3498459/>
- Andrillon, T., Nir, Y., Cirelli, C., Tononi, G., & Fried, I. (2020). Changes in cross-frequency coupling following closed-loop auditory stimulation in non-rapid eye movement sleep. *Scientific Reports*, 10, 10636. <https://pmc.ncbi.nlm.nih.gov/articles/PMC7326971/>
- Barakat, M., Doyon, J., Bherer, L., Fogel, S., Gagnon, J. F., & Carrier, J. (2011). Fast and slow spindle involvement in the consolidation of a new motor sequence. *Behavioural Brain Research*, 217(1), 117–121. <https://pubmed.ncbi.nlm.nih.gov/20974191/>
- Bendor, D., & Wilson, M. A. (2012). Biasing the content of hippocampal replay during sleep. *Nature Neuroscience*, 15, 1439–1444. <https://pubmed.ncbi.nlm.nih.gov/22922783/>
- Bernardi, N. F., De Buglio, M., Trimarchi, P. D., Chielli, A., & Bricolo, E. (2013). Mental practice promotes motor anticipation: Evidence from skilled music performance. *Frontiers in Human Neuroscience*, 7, 451. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3747442/>
- Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological Research*, 76(2), 192–203. <https://pubmed.ncbi.nlm.nih.gov/21541757/>
- Boutin, A., & Doyon, J. (2020). A sleep spindle framework for motor memory consolidation. *Philosophical Transactions of the Royal Society B*, 375(1799), 20190232. <https://pmc.ncbi.nlm.nih.gov/articles/PMC7209914/>
- Boutin, A., Pinsard, B., Bherer, L., & Doyon, J. (2023). Sleep-related motor skill consolidation and generalizability after physical practice, motor imagery, and action observation. *iScience*, 26(7). <https://pmc.ncbi.nlm.nih.gov/articles/PMC10374463/>
- Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature*, 382(6588), 252–255. <https://pubmed.ncbi.nlm.nih.gov/8717039/>
- Dave, A. S., & Margoliash, D. (2000). Song replay during sleep and computational rules for sensorimotor vocal learning. *Science*, 290(5492), 812–816. <https://pubmed.ncbi.nlm.nih.gov/11052946/>
- Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11(2), 114–126. <https://pubmed.ncbi.nlm.nih.gov/20046194/>
- Elmaleh, M., Kranz, D., Corredera Asensio, A., Elsherbiny, D., & Long, M. A. (2021). Sleep replay reveals premotor circuit structure for a skilled behavior. *Neuron*, 109(23), 3851–3861. <https://pmc.ncbi.nlm.nih.gov/articles/PMC8639717/>
- Ferreri, L., Mas-Herrero, E., Zatorre, R. J., Ripolles, P., Gomez-Andres, A., Alicart, H., & Rodriguez-Fornells, A. (2019). Dopamine modulates the reward experiences elicited by

- music. *Proceedings of the National Academy of Sciences*, 116(9), 3793–3798.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC6397525/>
- Fischer, S., Hallschmid, M., Elsner, A. L., & Born, J. (2002). Sleep forms memory for finger skills. *Proceedings of the National Academy of Sciences*, 99(18), 11987–11991.
<https://pubmed.ncbi.nlm.nih.gov/12011407/>
- Fogel, S. M., & Smith, C. T. (2011). The function of the sleep spindle: A physiological index of intelligence and a mechanism for sleep-dependent memory consolidation. *Neuroscience and Biobehavioral Reviews*, 35(5), 1154–1165.
<https://pubmed.ncbi.nlm.nih.gov/21167865/>
- Genzel, L., Kroes, M. C. W., Dresler, M., & Battaglia, F. P. (2015). The role of sleep in motor sequence consolidation: Stabilization rather than enhancement. *Journal of Neuroscience*, 35(17), 6696–6702. <https://pmc.ncbi.nlm.nih.gov/articles/PMC4412892/>
- Gold, B. P., Pearce, M. T., Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2023). Auditory and reward structures reflect the pleasure of musical expectancies during naturalistic listening. *Frontiers in Neuroscience*, 17, 1209398.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC10625409/>
- Haueisen, J., & Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, 13(6), 786–792.
<https://pubmed.ncbi.nlm.nih.gov/11564320/>
- Higgins, C., Liu, Y., Vidaurre, D., Kurth-Nelson, Z., Dolan, R., Behrens, T., & Woolrich, M. (2021). Replay bursts in humans coincide with activation of the default mode and parietal alpha networks. *Neuron*, 109(5), 882–893.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC7927915/>
- Hu, X., Cheng, L. Y., Chiu, M. H., & Paller, K. A. (2020). Promoting memory consolidation during sleep: A meta-analysis of targeted memory reactivation. *Psychological Bulletin*, 146(3), 218–244. <https://pmc.ncbi.nlm.nih.gov/articles/PMC7144680/>
- Joo, H. R., & Frank, L. M. (2018). The hippocampal sharp wave-ripple in memory retrieval for immediate use and consolidation. *Nature Reviews Neuroscience*, 19(12), 744–757.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC6794196/>
- Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience*, 22(10), 1598–1610.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC12576410/>
- Kuriyama, K., Stickgold, R., & Walker, M. P. (2004). Sleep-dependent learning and motor-skill complexity. *Learning and Memory*, 11(6), 705–713.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC534699/>
- Lewis, P. A., & Durrant, S. J. (2011). Keeping time in your sleep: Overnight consolidation of temporal rhythm. *Neuropsychologia*, 49(1), 115–123.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC7614423/>
- Manoach, D. S., Stickgold, R., Merrill, C., Mikulka, E., Armstrong, J., Jain, A., & Garrett, A. (2024). The effects of closed-loop auditory stimulation on sleep oscillatory dynamics in

- relation to motor procedural memory consolidation. *Sleep*, 47(4).
<https://pmc.ncbi.nlm.nih.gov/articles/PMC11009689/>
- Margoliash, D., Olshe, A., Dave, A. S., & Yu, A. C. (2001). Dynamic control of auditory activity during sleep: Correlation between song response and EEG. *Proceedings of the National Academy of Sciences*, 98(24), 14012–14016.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC61158/>
- Nishida, M., & Walker, M. P. (2007). Daytime naps, motor memory consolidation and regionally specific sleep spindles. *PLoS ONE*, 2(4), e341.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC1828623/>
- Palmer, C., & Mathews, S. (2013). Auditory and motor imagery modulate learning in music performance. *Frontiers in Human Neuroscience*, 7.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC3696840/>
- Pranjić, M., Braun Janzen, T., Vukšić, N., & Thaut, M. (2024). From sound to movement: Mapping the neural mechanisms of auditory–motor entrainment and synchronization. *Brain Sciences*, 14(11), 1063. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11592450/>
- Rasch, B., Büchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science*, 315(5817), 1426–1429.
<https://pubmed.ncbi.nlm.nih.gov/17347444/>
- Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews*, 93(2), 681–766. <https://pubmed.ncbi.nlm.nih.gov/23589831/>
- Rakowska, M., Bagrowska, P., Lazari, A., Navarrete, M., Abdellahi, M. E. A., Johansen-Berg, H., & Lewis, P. A. (2025). Targeted memory reactivation elicits temporally structured reactivation linked to spindles. *Imaging Neuroscience*. Advance online publication.
<https://direct.mit.edu/imag/article/doi/10.1162/IMAG.a.1123/134892/>
- Robertson, E. M., Pascual-Leone, A., & Press, D. Z. (2004). Awareness modifies the skill-learning benefits of sleep. *Current Biology*, 14(3), 208–212.
<https://pubmed.ncbi.nlm.nih.gov/14738728/>
- Robertson, E. M., Albouy, G., & Bhatt, N. (2018). Dual enhancement mechanisms for overnight motor memory consolidation. *Nature Human Behaviour*, 2, 160–166.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC5839513/>
- Rudoy, J. D., Voss, J. L., Westerberg, C. E., & Paller, K. A. (2009). Strengthening individual memories by reactivating them during sleep. *Science*, 326(5956), 1079.
<https://pubmed.ncbi.nlm.nih.gov/19892986/>
- Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature Neuroscience*, 14(2), 257–262. <https://pubmed.ncbi.nlm.nih.gov/21217764/>
- Schlaug, G. (2015). Musicians and music making as a model for the study of brain plasticity. *Progress in Brain Research*, 217, 37–55.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC4430083/>

- Schönauer, M., Geisler, T., & Gais, S. (2014). Strengthening procedural memories by reactivation in sleep. *Journal of Cognitive Neuroscience*, 26(1), 143–153.
<https://pubmed.ncbi.nlm.nih.gov/24036511/>
- Simmons, A. L. (2012). Effects of practice variability and distribution of practice on musicians' performance of a procedural skill. *Journal of Research in Music Education*, 59(4), 357–368.
- Simmons, A. L., & Duke, R. A. (2012). Distributed practice and procedural memory consolidation in musicians' skill learning. *Journal of Research in Music Education*, 59(4), 321–336.
- Stickgold, R. (2000). Visual discrimination learning requires sleep after training. *Nature Neuroscience*, 3(12), 1237–1238. <https://pubmed.ncbi.nlm.nih.gov/11100141/>
- Tucker, M. A., Nguyen, N. D., & Stickgold, R. (2016). Experience playing a musical instrument and overnight sleep enhance performance on a sequential typing task. *PLoS ONE*, 11(7), e0159608. <https://pmc.ncbi.nlm.nih.gov/articles/PMC4966959/>
- Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with sleep makes perfect: Sleep-dependent motor skill learning. *Neuron*, 35(1), 205–211.
<https://pubmed.ncbi.nlm.nih.gov/12123620/>
- Walker, M. P. (2005). Sleep and the time course of motor skill learning. *Learning and Memory*, 12(3), 229–237. <https://pmc.ncbi.nlm.nih.gov/articles/PMC202318/>
- Wan, C. Y., & Schlaug, G. (2010). Music making as a tool for promoting brain plasticity across the lifespan. *Neuroscientist*, 16(5), 566–577.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC4430083/>
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8(7), 547–558. <https://pubmed.ncbi.nlm.nih.gov/17585307/>
- Zatorre, R. J., & Salimpoor, V. N. (2013). From perception to pleasure: Music and its neural substrates. *Proceedings of the National Academy of Sciences*, 110(Suppl 2), 10430–10437. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3690607/>

The Determinant Indicator Of Climate Risk By Frank Ma

Abstract

The paper investigates the determinant factor of climate risk, focusing on how socio-economic indicators influence environmental vulnerability. Utilising a comprehensive dataset covering 239 countries, the study examines the relationship between climate risk and socio-economic variables. The empirical results demonstrate that government effectiveness, forest area, inflation, labour force participation, and research and development (R&D) expenditure have a negative correlation with climate risk. Conversely, GDP growth is found to have a positive effect on climate risk. The paper provides critical evidence for policy makers and organisations, suggesting ways to reduce environmental vulnerability through socio-economic development.

Keywords Economics, Macroeconomics, Climate Risk, Inflation, Labour Force, GDP Growth

Introduction

The study examines how socio-economic factors influence a country's vulnerability to environmental hazards. While, climate risk is often viewed solely as an environmental problem, it is fundamentally shaped by underlying economic and social conditions.

Recent academic studies have extensively explored the effect of climate risk on economic growth, labour productivity and inflation. However, this study argues that climate risk is influenced not only by environmental factors but also by institution, economic growth and social variables.

The purpose of the study is to investigate the relationship between climate risk and socio-economic indicators using scatter plot analysis. The findings indicate that government effectiveness, forest area, inflation, labour force, research and development (R&D) expenditure are negatively correlated with climate risk; while GDP growth shows a positive correlation with climate risk.

While much of existing literature focuses on environmental factors, this paper fills the gap in the field by highlighting socio-economic determinants. Using 239 countries, the study offers robust empirical evidence.

The paper is organized as follows: Section 2 provides background information on the existing literature, Section3 highlights the empirical findings, and Section 4 concludes the paper.

Litureture Review

Various methodologies have been employed to investigate the determinant indicator of climate risk, including OLS (Halkos and Zisiadou, 2023), General Equilibrium Model (Komendantova et al., 2019), Meta analysis (Dasgupta and Robinson, 2023), Theoretical model (Alexandria et al., 2024), Granger causality test (Zhang, 2023), Descriptive statistics (Nathwani

et al., 2021; Rod et al., 2025), NPP (Yin et al., 2018), LA-VAR (Zhang, 2023), GMM Method (Qi et al., 2025).

The variable used in the literature is a determinant factor of climate risk varied such as: IPCC (Rod et al., 2025), Labour productivity (Dasgupta and Robinson, 2023), FDI (Komendatova et al., 2019), Labour force (Dasgupta and Robinson, 2023), Employment (Alexandria et al., 2024), Inflation (Zhang, 2023), population (Halkos and Zisadou, 2023), Human development index (Nathwani et al., 2021), Forest area (Yin et al., 2018),

Many studies have focused on various countries and regions such as: Global (Halkos and Zisiadou, 2023; Nathwani et al., 2021), MENA (Komendatova et al., 2019), Europe (Alexandria et al., 2024), BRICS (Zhang, 2023), China (Yin et al., 2018), Sub-Saharan Africa (Dasgupta and Robinson, 2023), Norwegian (Rod et al., 2025).

The economic consequences of climate risk have received growing attention in recent years, examining its impact through various channels such as economic growth, labour productivity and inflation. Many papers focusing on the relationship between the climate risk indices and social economic performance of countries for example, Halkos and Zisiadou (2023) analyze the relationship between climate change performance index and climate risk index, their findings indicate non statistically significant relationship between two indices. However they emphasise the role of economic growth, population dynamics and excess to electricity in shaping these indices. Additionally, Nathwani et al. (2021) propose the development and climate risk change performance index. Her important literature investigates the macroeconomic implications of climate change and climate risk, for example Alexandria et al. (2024) employ macroeconomic models for euro and show that climate policies have significant effect on GDP and employment. Similarly Komendantova et al. (2019) use the CGE model and demonstrate that sustainability has significantly effectives on economic performance.

Table 1. An extensive review of Literature on Determinant Indicator of Climate Risk

Authors	Year	Variables	Methodology	Region/ Country	Findings	Period
Halkos & Zisiadou	2023	CRI, GDP per capita, Access to Electricity, Population Density	OLS	EU	No direct link between risk (CRI) and climate action (CCPI)	2019
Komendantova et al.	2019	Labour Capital, FDI, Investment Cost	General Equilibrium (CGE) model	MENA	De-risking strategies can boost foreign investment, lowering costs	2004 -2009
Dasgupta & Robinson	2023	Labour Force, Economic Growth, Workers Health	Literature Review	Global	Rising heat significantly reduces outdoor labor productivity, worsening inequality	2019 - 2022

Alexandri et al	2024	GDP, Employment	Multi Sectorial Macroeconomic Model	EU	Mitigation policies can modestly boost jobs/GDP while reducing labour market inequality.	2015 - 2030
Zhang	2023	Inflation	Granger Causality Test	BRICS	Climate risk influences inflation in several BRICS countries	1992 - 2007
Nathwani et al.	2021	DCI, Human Development Index, CCPI	Descriptive Statistics	Global	The paper presents the Development and Climate Change Performance Index (DCI)	2019
Yin et al.	2018	Forest Area	NPP	China	Climate change will reduce forest productivity in China,	2009-2013
Qi et al.	2025	CPI, Government Debt, GDP, M2	GMM	Global	Climate change raises inflation in the short term, with stronger effects in low-income countries.	2006-2019
Rod et al.	2025	IPCC, Hazard, Vulnerability	Descriptive statistics	Norwegian	A four-factor indicator framework helps assess municipalities' climate adaptation needs and performance.	1980-2023

Empirical Evidence

Data

The indicators used in this study to determine climate risk include government effectiveness, forest area as a percentage of land area, inflation as a consumer price, labour force participation rate as a percentage of total population ages 15-64. The data is sourced from the World Bank World Development Indicator, utilizing the most recent available dataset from 2023.

Descriptive Statistics

Table1. Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Climate risk	239	193.76	308.1	-82.39	1856
Gov. Effectiveness	185	-.045	.985	-2.233	2.144
GDP Growth	233	4.401	6.088	-28.758	63.334
Forest Area	235	31.87	22.18	0	94.524
Inflation	209	11.728	17.713	1.045	171.205
Labor Force	223	66.553	11.734	23.615	88.527

Table 1 summarizes the descriptive statistics for variables analyzed. For example, the climate risk has 239 observations with a mean of 193.76 and standard deviation of 308.1, ranging from a minimum of -82.39 to maximum of 1856. The government effectiveness has 185 observations with a mean of -0.45 and standard deviation of 0.985, ranging from a minimum of -2.233 to maximum 2.1444. The GDP growth has 233 observations with a mean of 4.401 and standard deviation of 6.088, ranging from a minimum of -28.758 to a maximum of 63.334. The forest area (percentage of land area) has 235 observations with a mean of 31.87 and standard deviation of 22.18, ranging from a minimum 0 to a maximum of 94524. The inflation has 209 observations with a mean of 11.728 and standard deviation of 17.713, ranging from a minimum 1.045 to a maximum 171.205. The labour force participation (percentage of population rates) has 223 observations with a mean of 66.553 and standard deviation of 11.734, ranging from a minimum 23.615 to a maximum 88.527.

Empirical Findings

Figure 1: Scatter plot. Graphic of Climate Risk and Government Effectiveness

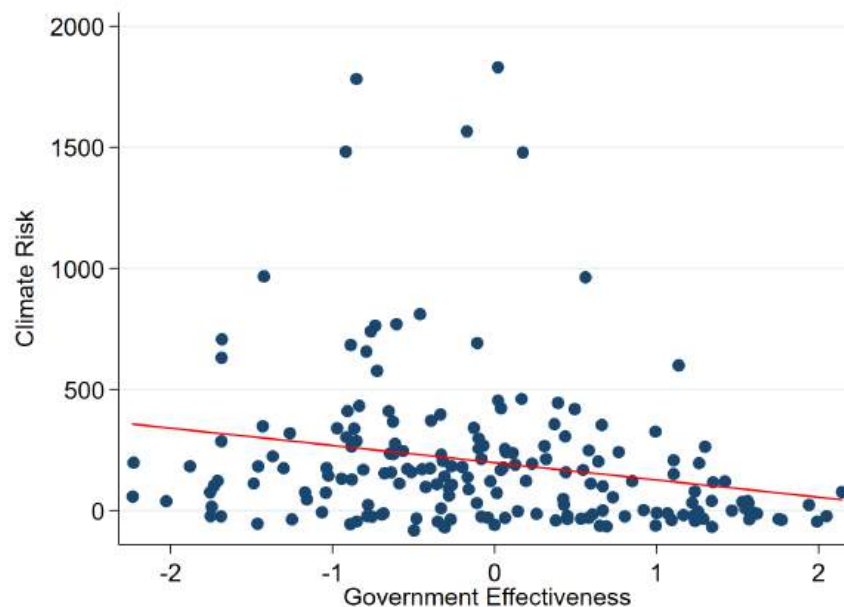


Figure 1 shows the connection between climate risk and government effectiveness. It is clear that government effectiveness negatively correlates with climate risk suggesting that higher level of government effectiveness contributes to lower level of climate risk.

Figure 2: Scatter plot, Graphic of climate risk and GDP growth

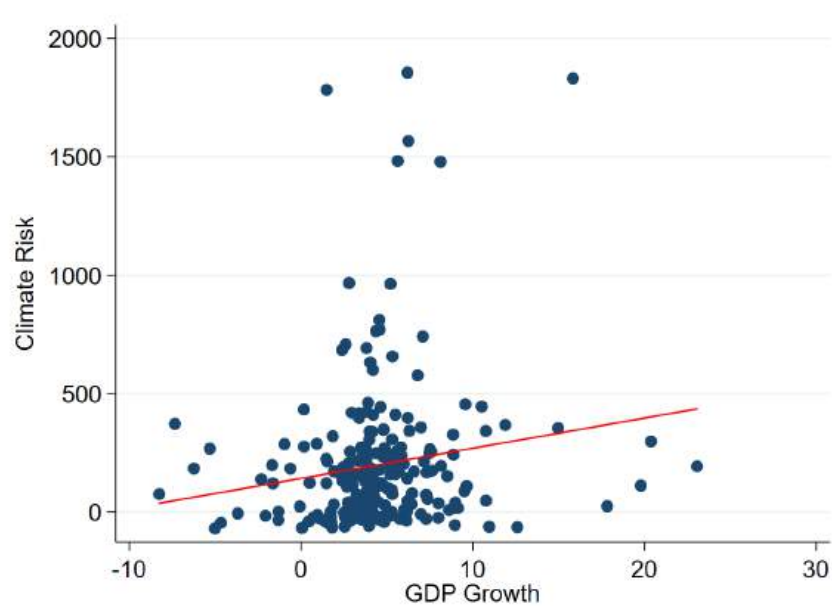


Figure 2 shows the relationship between climate risk and GDP growth. It is clear that GDP growth positively correlates with climate risk suggesting that higher levels of GDP growth tend to be higher levels of climate risk. Some fast growth economies depend on industry and manufacturing that contribute to climate risk.

Figure 3 : Scatter plot. Graphic of Climate Risk to forest area

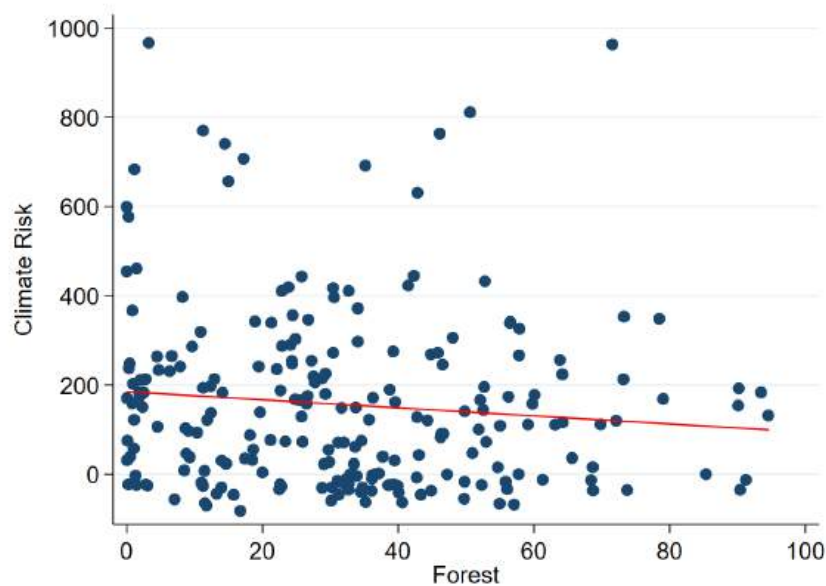


Figure 3 shows the relationship between climate risk and the forest area. It is clear that forest area negatively correlates with climate risk suggesting that higher levels of forest area tend to be lower levels of climate risk.

Figure 4: Scatter plot, Graphic of climate risk and inflation

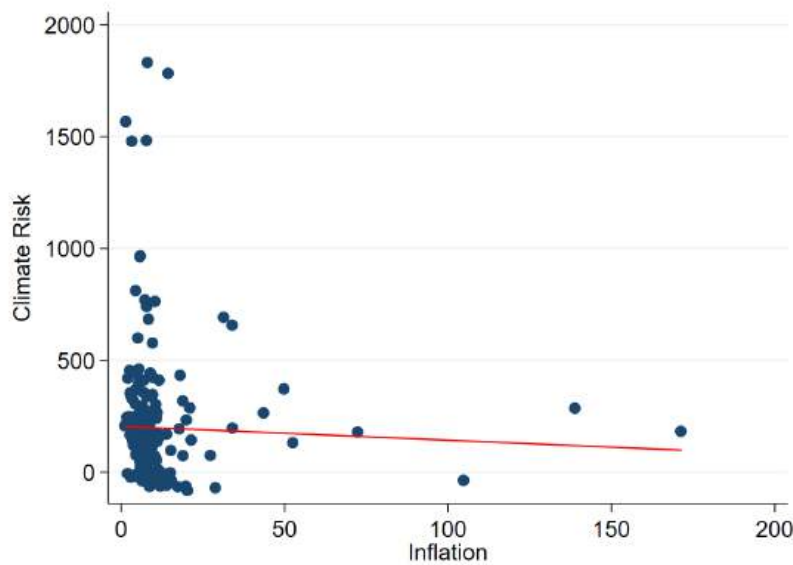


Figure 4 shows the relationship between climate risk and inflation. It is clear that inflation slightly negatively correlates with climate risk suggesting that higher levels of inflation tend to be lower levels of climate risk. It is clear that when inflation is high people and governments focus on immediate economic problems and spend less on long term climate solutions, increasing climate risk.

Figure 5: Scatter plot, Graphic of climate risk and labour force

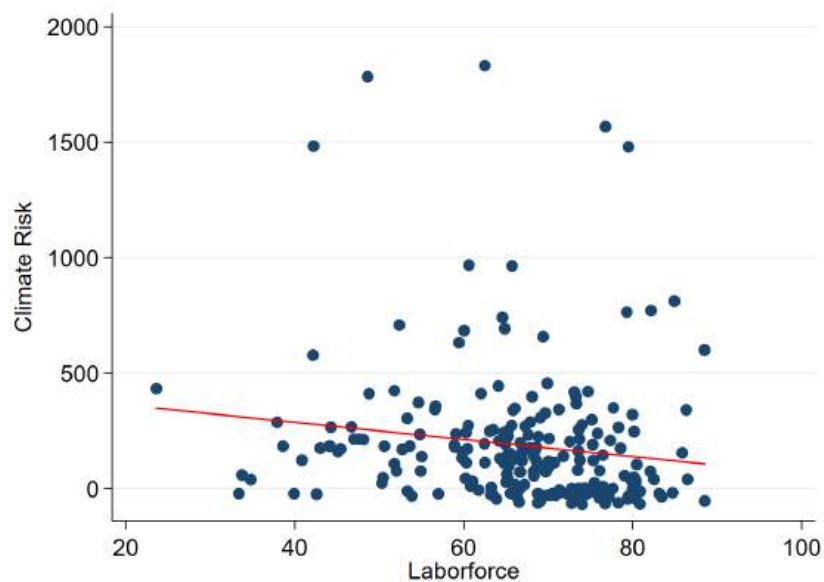


Figure 5 shows the relationship between climate risk and labour force participation rate. It is clear that inflation negatively correlates with climate risk suggesting that as climate risk increases, fewer people are participating in the labour force.

Figure 6: Scatter plot, Graphic of climate risk and research and development expenditure as a % of GDP

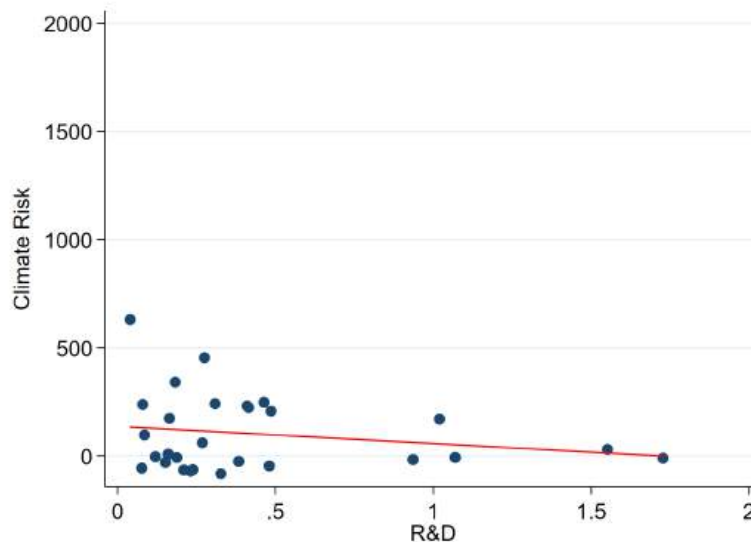


Figure 6 shows the relationship between climate risk and research and development expenditure as a % of GDP. It is clear that research and development expenditure as a % of GDP slightly negatively correlates with climate risk suggesting that higher spending on research and development leads to better technology and solutions for climate adaptation and mitigation which help reduce climate risk.

Conclusion

This paper investigated the determinant of climate risk by analysing the empirical relationship between environmental vulnerability and socio-economic indicators. The result confirms the negative correlation between government effectiveness and climate risk, suggesting that high quality public services are critical in reducing environmental vulnerability. Similarly, higher expenditure on Research and Development (R&D) contributes to more efficient climate adoption strategies and innovative technological solutions.

The study also shows a positive correlation between GDP growth and climate risk, which is likely derived by the dominance of industrial and manufacturing sectors in rapidly growing economies. Furthermore, climate risk appears to have a significant implication in the labour force market; the findings suggest that the higher level of labour force links to a lower level of climate risk. Additionally inflation shows slightly negative correlation with climate risk. High inflation often forces the government to take into account climate issues.

The study highlights that climate risk is not merely an environmental issue but is deeply embedded in a country's socio-economic framework.

Future study should employ longer time series data and panel data analysis to capture more robust evidence.

Works Cited

- Kouloukoui, D., da Silva Gomes, S. M., Torres, F. A., & Torres, E. A. (2025). Business climate risk management: international perspectives and strategic determinants. *Environment, Development and Sustainability*, 27(2), 4683-4724.
- Komendantova, N., Schinko, T., & Patt, A. (2019). De-risking policies as a substantial determinant of climate change mitigation costs in developing countries: Case study of the Middle East and North African region. *Energy Policy*, 127, 404-411.
- Dasgupta, S., & Robinson, E. J. (2023). The labour force in a changing climate: research and policy needs. *PLOS Climate*, 2(1), e0000131.
- Dasgupta, S., & Robinson, E. J. (2023). The labour force in a changing climate: research and policy needs. *PLOS Climate*, 2(1), e0000131.
- Alexandri, E., Antón, J. I., & Lewney, R. (2024). The impact of climate change mitigation policies on European labour markets. *Ecological Economics*, 216, 108022.
- Zhang, Z. (2023). Are climate risks helpful for understanding inflation in BRICS countries?. *Finance Research Letters*, 58, 104441.
- Halkos, G., & Zisiadou, A. (2023). The effects of climate change to weather-related environmental hazards: interlinkages of economic factors and climate risk. *Journal of Risk and Financial Management*, 16(5), 264.
- Nathwani, J., Lind, N., Renn, O., & Schellnhuber, H. J. (2021). Balancing health, economy and climate risk in a multi-crisis. *Energies*, 14(14), 4067.
- Yin, Y., Ma, D., & Wu, S. (2018). Climate change risk to forests in China associated with warming. *Scientific reports*, 8(1), 493.
- Zhang, Z. (2023). Are climate risks helpful for understanding inflation in BRICS countries?. *Finance Research Letters*, 58, 104441.
- Qi, C., Ma, Y., Du, M., Ma, X., Xu, Y., & Zhou, X. (2025). Impacts of climate change on inflation: An analysis based on long and short term effects and pass-through mechanisms. *International Review of Economics & Finance*, 98, 103846.
- Dasgupta, S., & Robinson, E. J. (2023). The labour force in a changing climate: research and policy needs. *PLOS Climate*, 2(1), e0000131.
- Rød, J. K., Aall, C., & Selseng, T. (2025). Towards a holistic climate service: Addressing all four climate risk determinants. *Climate Services*, 38, 100558.

Appendix

Table 2. Definition of Variables

Sembol	Variable	Definition	Source
Climate Risk	Carbon dioxide (CO2) emissions (total) excluding LULUCF (% change from 1990)	Change of emissions (as %) of current year with respect to emissions in baseline year 1990 emissions of carbon dioxide (CO2), one of the six Kyoto greenhouse gases (GHG), from the agriculture, energy, waste, and industrial sectors, excluding LULUCF	WDI: World Bank

Government effectiveness	Government effectiveness: Estimate	Government Effectiveness captures perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures.	WDI: World Bank
Agricultural Land	Agricultural land (% of land area)	Agricultural land refers to the share of land area that is arable, under permanent crops, and under permanent pastures.	WDI: World Bank
Forest Area	Forest area (% of land area)	Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems	WDI: World Bank
Inflation	Inflation, consumer prices (annual %)	Inflation as measured by the consumer price index reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services	WDI: World Bank
Labour Force Participation	Labor force participation rate, total (% of total population ages 15-64) (modeled ILO estimate)	Labor force participation rate is the proportion of the population ages 15-64 that is economically active.	WDI: World Bank

The Relationship Between Female Political Representation and Sports Development in India By Anaya Tahilramani

Abstract

Over two decades beginning in 2000, patterns linking women's presence in national politics and progress in sports within India are explored here. Rather than focusing solely on numbers, shifts in legislative participation by females take center stage alongside financial allocations directed toward athletic programs. Information stems from official budget records, nationwide electoral reports, and global economic databases. Statistical techniques gauge how closely changes in one domain align with developments in another. While GDP per capita serves as an indicator of broader prosperity, its interplay with policy-related inputs is examined without assuming causation. Findings emerge through numerical relationships drawn from structured datasets gathered at regular intervals. Though correlation does not confirm influence, repeated observations allow for scrutiny of long-term tendencies. Although women's presence in the Lok Sabha rose slowly during the observation window, it continues to trail behind worldwide levels, dropping again in the latest poll. Alongside broader financial growth, money set aside for sports has risen sharply. When taken together, the share of women in national politics and overall economic output account for much of the shift seen in athletic financing. Despite gradual gains, gender balance in legislative roles stays limited when measured internationally. As the country moves ahead with its new law reserving seats for women, deeper inquiry could uncover why such patterns emerge. Methods focusing on personal insight may reveal what drives this connection over time.

Keywords Women, Representation, Sports, Development, Politics, GDP

1. Introduction

Despite notable achievements by female athletes internationally, along with growing worldwide dialogue on fairness between genders, gaps remain pronounced within sports for women in India. Rooted deeply, these imbalances emerge through tangled networks of social norms, governance patterns, and financial realities. The argument here suggests such inequality in athletic opportunity does not stand apart - instead it mirrors wider national frameworks, particularly how women engage politically and how resources are shaped across the economy.

One way to examine this issue involves looking at how women's influence in politics might shape sports growth in India. The share of women in the Lok Sabha offers a clear indicator of their role in shaping policy decisions. Where legislative bodies include more women, research suggests greater attention often goes toward supporting female athletes through public funding or fair compensation measures. When it comes to India, participation by women in parliamentary seats has stayed minimal throughout recent decades, possibly reducing momentum behind efforts to advance women's involvement in competitive sport.

Female athletes in India show contrasting realities when observed closely. While global competitions highlight achievements - Olympic podiums, the hockey squad reaching a rare semifinal, consistent results at the Commonwealth level - a different pattern emerges closer to home. Participation among young girls in schools remains limited, data suggests. Far fewer take part in structured physical activity when measured against male peers. Recognition abroad does not mirror engagement within communities. Teenage years bring a steep rise in school dropouts among girls, driven by inadequate infrastructure, worries about personal security, along with societal norms shaping their roles. Although high-achieving women in athletics may succeed, deep-rooted obstacles still block wider access for females in sport.

Economic standing, typically reflected through Gross Domestic Product (GDP), influences how nations support athletic development. Though India has seen notable GDP growth across twenty years, enabling expanded government spending on sport, wealth alone does not ensure balanced access. Decisions about where money goes depend on policy choices - persistent societal preferences frequently prioritize male athletics. Even with larger total investments in sports lately, financial support aimed at female-focused initiatives and facilities lacks consistent measurement or transparency.

Beginning with India, attention turns to structural factors shaping women's sports rather than simple contrasts. Instead, numbers guide the way - linking female presence in governance, financial capacity of the nation, funding dedicated to athletics, and actual involvement of women in competitive events. What emerges sits beneath the visible: performance arenas mirror long-standing choices made in policy and economy when it comes to sex-based fairness. Outcomes are less about games, more about where society places value.

2. Methodology

This methodology explains the process used to explore whether there is a connection between the number of women in India's parliament and the funding allocated to sports development. The study looks at data over a long period, from 2000 to 2024. The main challenge was finding data that was both reliable and comparable across different years and sources. This section describes how that was done and why certain methods were chosen. It also explains the steps taken to analyze the data and the limitations of this approach.

2.1 Research Aim

The main goal was straightforward but challenging: to determine whether gains in women's political power at the national level in India coincided with more government money being directed toward sports. This idea came from reading other studies that suggest diverse legislatures sometimes fund different priorities. To test this, several smaller questions needed to be answered first.

First, the basic trends had to be tracked. How has India's economy grown? How has women's presence in the Lok Sabha changed over the years? And how had sports spending

moved during the same time? Simply plotting these on a graph provides a first look at any possible relationship.

Second, the study aimed to move beyond simple observation. To do this, statistics were used to measure how closely these factors moved together. The question was whether increases in women's representation were statistically linked to increases in sports budgets, even when considering general economic growth.

Finally, numbers only tell part of the story. A key aim was to interpret the statistical results within India's unique political context. For instance, a significant increase in women's representation in 2019 needed to be discussed in light of that election. The analysis had to blend the quantitative findings with qualitative, real-world explanations.

2.2 Research Design

Given the nature of the research question, which is about change over time, a longitudinal design was the obvious choice. This simply means we collected data from many different points in time to see how things evolved. We decided to use a quantitative approach because we were dealing with budgets and seat counts, which are easily measured in numbers. This type of design is common in political science and economics when researchers are studying policy outputs.

One option might have involved examining a single state in depth or conducting multiple discussions with elected officials. Still, such approaches could not deliver the wide-reaching, countrywide perspective aimed for here. What supports the selected approach is its ability to uncover trends across vast data sets. Yet a limitation remains evident: personal choices or bargaining processes within politics stay hidden. The presence of connections may appear clearly, though their inner function does not.

2.3 Data Collection

The research paper relies exclusively upon publicly accessible data from reliable official sources to establish its credibility.

World Bank statistics regarding India's annual Gross Domestic Product (GDP) growth and GDP per person provided the basis for measuring India's economic health. The use of these World Bank statistics ensured that there would be consistency in the measurements being applied.

Calculating the extent of women's political participation was relatively straightforward. Each post-election, the number of women elected to the Lok Sabha was divided by the total number of Lok Sabha representatives. This data was obtained from the official statistical records of the Election Commission of India. To validate and expand the time frame of this data set, it was compared against the World Bank's Gender Statistics Database, which aggregates the same percentages.

Assembling data related to sports funding presented perhaps the largest challenge. No single ideal source existed; therefore, all of the figures had to be gathered manually from World Bank Data and Statista. Only an approximate value in billions of Rupees for sports expenditures

through the central government could be identified each year. Reliable and consistent data in this format became available, which also marked the beginning date for a comprehensive statistical analysis.

2.4 Organizing and Preparing the Data

Before any analysis could begin, the raw data needed to be organized and adjusted. All the figures, economic, political, and budgetary, were entered into a single master spreadsheet sorted by year. This allowed for easy comparison.

A technical but important step involved the financial data. Budgets and economic output tend to grow multiplicatively (e.g., increasing by a percentage each year), not simply additively. This can distort statistical tests. To correct for this, we applied a natural logarithm transformation to the sports expenditure and GDP per capita figures. In simpler terms, we converted them to a scale better suited to the linear analysis we planned to do. These transformed variables are labeled as “log expenses” and “log GDP” in the results.

2.5 Data Analysis

The analysis happened in two distinct phases. The first phase was about seeing the story. We created basic line charts to visually trace the path of GDP growth and women’s representation from 2000 onward. Looking at these graphs, we can spot major turning points, such as a dip after a global crisis or a spike after an election. We annotated these charts with brief notes about historical events. This visual stage is crucial. It prevents researchers from jumping straight into complex statistics without first understanding the raw shape of the data.

The second phase was the statistical testing, focused on the 2013-2024 period, where we had complete data. We started with a correlation analysis. This creates a matrix that shows, with a number between -1 and 1, how strongly two variables are related. For example, it tells us if years with higher women’s representation also tend to be years with higher sports spending.

Correlation is useful, but it doesn’t control for other factors. The next step was regression analysis. This more powerful technique allowed us to ask: “Can we predict sports spending based on both GDP and women’s representation?” We set up a model where sports spending (log expenses) was the “outcome” we wanted to predict. The “predictors” were the country’s wealth (log GDP) and the percentage of women in parliament. The software (Google Sheets) then calculated how much of the variation in sports spending was explained by these two factors together, producing an R-squared value and a significance test (p-value).

2.6 Ethical Stance

Since this study works only with published national statistics, there were no ethical concerns involving human subjects, privacy, or confidential information. All data is in the public domain.

The primary ethical consideration here is one of intellectual honesty. We must be clear about what the data can and cannot tell us.

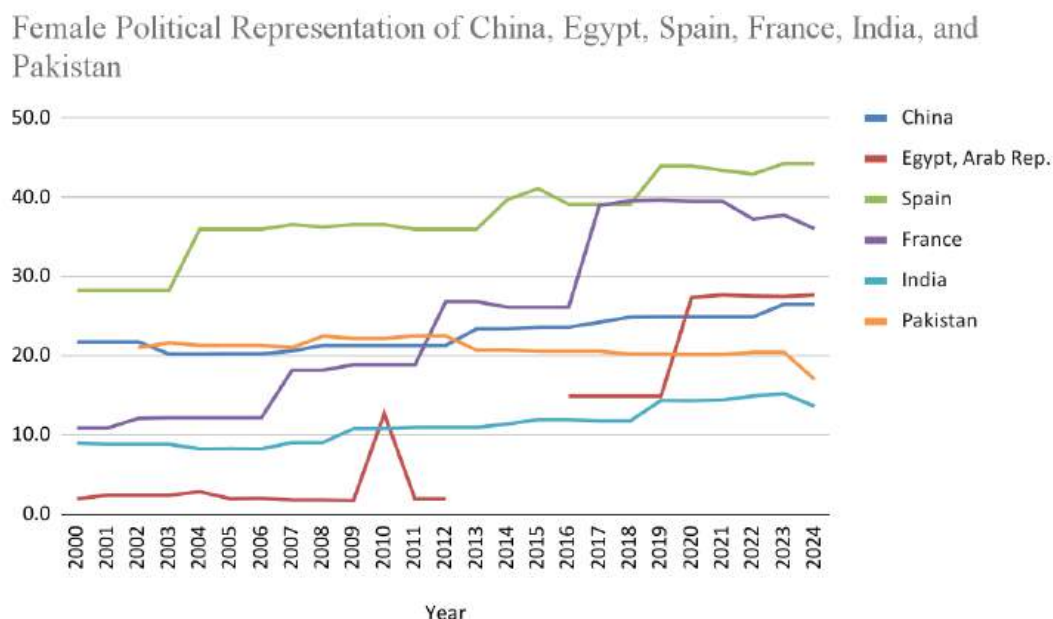
2.7 Limitations

The most important limitation is that correlation does not prove causation. Just because women's representation and sports spending rise together does not mean one causes the other. There could be a third factor, like a general national focus on development, that drives both. The regression model can control for some factors like overall wealth, but it cannot control for everything.

Furthermore, the sports budget is a very broad measure. It does not tell us how the money was spent, or whether it reached grassroots programs for women and girls. A future study might look at that. This research is a necessary first step: it establishes whether a noteworthy pattern exists at the macro level that deserves further, more detailed investigation. We present the findings with these caveats firmly in mind, aiming to contribute to a discussion rather than offer a definitive answer.

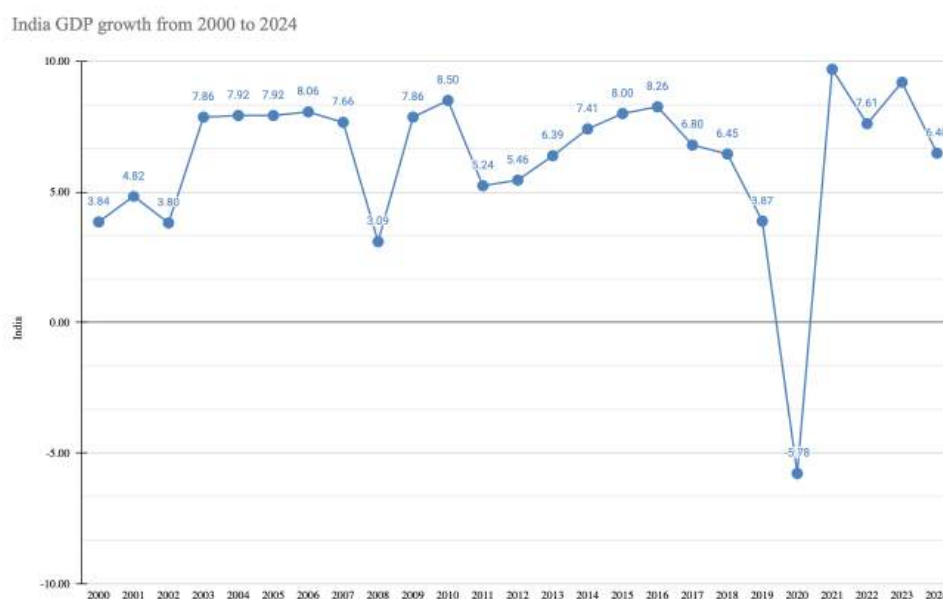
Results and Discussion

This section reports the results of an investigation of the relationship between female political representation and sports development in India. It examines patterns in legislative gender parity and their association with athletic infrastructure and participation rates, followed by a critical discussion of the implications in light of existing literature. The findings are presented through both descriptive statistics and visual representations, including comparative graphs of regional development data.



Graph 1: Female Political Representation of China, Egypt, Spain, France, India and Pakistan from 2000 to 2024 (World Bank Data)

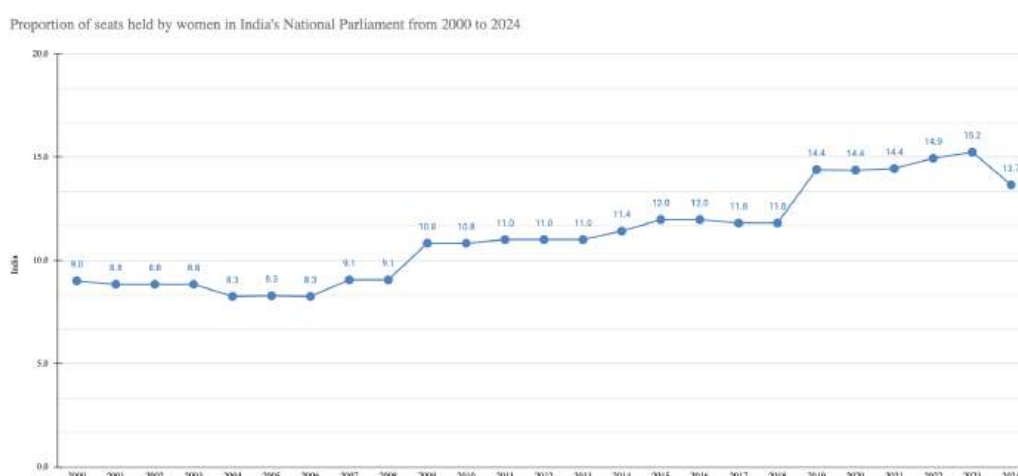
In 2024, China was 6.7 percent higher than it had been in 2000 at 21.8% and reached 26.5%. Egypt began at an all time low of just 2%, however, made a rapid jump to 27.7% in 2024. Spain went from 28.3% to 44.3% over the course of this period; it remained the highest. France also experienced a significant increase in its rate going from 10.9% to 36%, with some spikes occurring in 2007, 2012, and 2017. The rate in India increased by less than one percentage point per year, increasing from 9% to a high of 15.2% in 2023, and then decreasing to 13.7% in 2024. Although both Pakistan's rates were generally steady around the range of 20-22% each year, the country suffered a sharp decline in 2024 to 17%. As such, Spain and China have shown the greatest overall increases, while India and Pakistan have seen some declines recently.



Graph 2: GDP Growth in India From 2000 to 2024 (World Bank Data)

India has maintained one of the world's fastest growth rates over the last two decades (International Monetary Fund, 2024; Reserve Bank of India, 2024). Despite a significant decline in the 2020 pandemic. Between 2000 and 2002, growth was subdued (between 3.8% and 4.8%), as this period was considered a recovery from the Asian Financial Crisis of 1997 and was negatively influenced by the worldwide "Dot-com crash" (Frontline The Hindu, 2000) and drought conditions (BBC News, 2002) that prevailed in some agricultural areas of India. During the period from 2003 until 2007, the Indian Economy accelerated and reached sustained high growth, averaging almost 8 percent annually (Bräuning, 2025). The driving force behind this growth was the global commodity boom, increased exports of software and IT services, and an unprecedented rate of infrastructure development (e.g., Golden Quadrilateral project). In 2008, growth declined sharply to 3.09% due to the global liquidity crisis (The Economist, 2008). In 2009 (7.86%) and 2010 (8.5%), India bounced back with a textbook-style recovery, driven by fiscal stimulus packages rolled out by the Government of India. In 2011, India's growth again

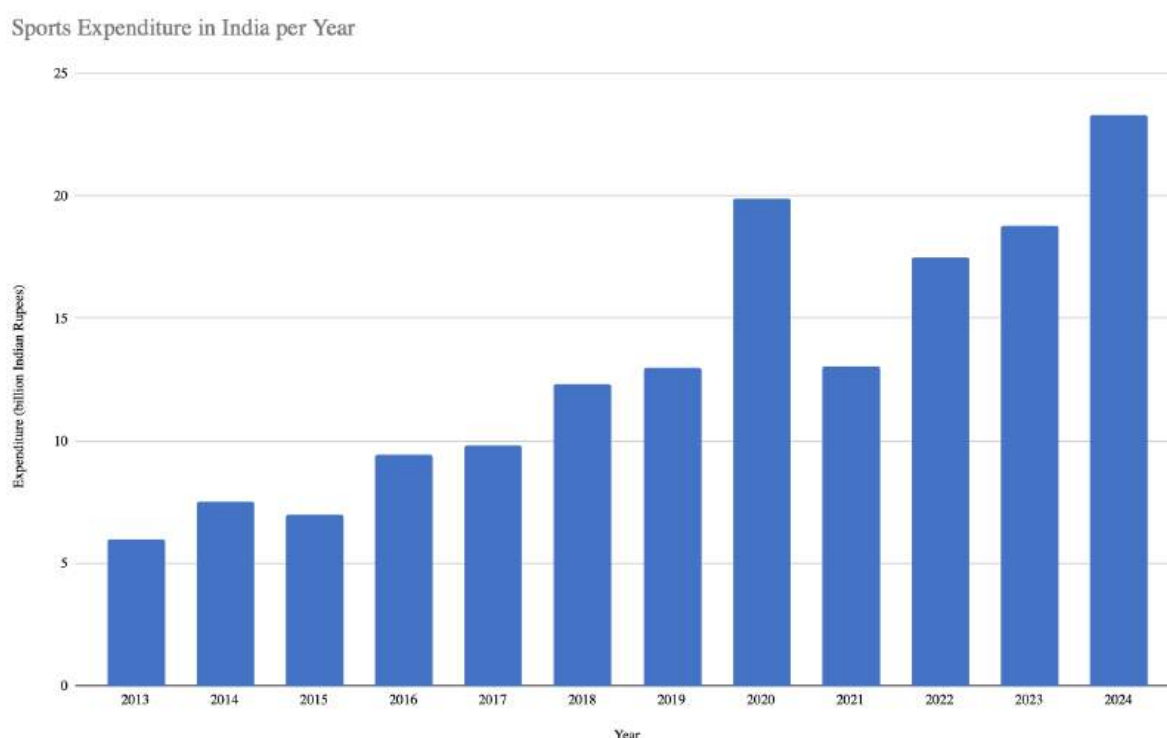
bounced back to 5.24% due to rising inflation and interest rates, which limited growth. This was followed by Reform policies and a decline in oil prices, bringing growth back steadily to 8.26% by 2016. In 2019, growth decreased again to 3.87%. This pre-pandemic slowdown is sometimes referred to as the "Twin Balance Sheet" problem (bad loans in banks) and the transitional adjustment caused by Demonetization (late 2016) and the GST (Goods and Services Tax in 2017) (Subramanian & Felman, 2019; Pandey & Sengupta, 2016; Press Trust of India, 2017). In 2020, India saw its first-ever economic contraction (-5.78%) due to nationwide lockdowns that led to a complete shutdown of the economy. Following this contraction was a base-effect bounce back of 9.7% as the economy reopened in 2021. Since 2022, India's economy has stabilised in a new growth range of 7-8%. Despite the global headwinds India faces (the Ukraine-Russia Conflict, Global Inflation), India has remained the fastest-growing major economy on the planet.



Graph 3: Proportion of Seats Held by Women in India's National Parliament from 2000 to 2024

In the early 2000s, women's representation in political office was below 9%. Even though women's success in Panchayati Raj (local government) was driven by the 73rd and 74th Amendments (PRS Legislative Research, 2023) to the Constitution, this "grassroots surge" did not immediately translate into success at the national level in Indian politics. After declining to 8.3% in 2004, it became evident that major political parties had established a "glass ceiling" in candidate selection. The 2009 election was a psychological turning point for women's representation, as the proportion of women in political office crossed the 10% threshold for the first time. Women began gaining greater visibility in high-profile roles at this time, including the Speaker of the Lok Sabha and the Leader of the Opposition, thereby normalizing women's leadership in the national consciousness. The 14.4% representation in 2019 was the highest level of women's representation in India since independence. The 2024 elections did see a modest decline in female representation to 13.6% (74). Women's representation in India has increased from less than 9% in 2000 to nearly 14% in 2024, but it remains significantly below the global

average of ~26.9%. The end of this gradual period from 2000–2024 will mark the transition from "voluntary participation" to "statutory reservation".



Graph 4: Sports Expenditure in India per Year from 2013 to 2024 (World Bank Data)

Starting at 6.5 billion rupees in 2013, sports spending in India climbed to 23.0 billion by 2024 - about three and a half times higher. For two years after 2013, outlays hovered near 7 billion, showing little movement. From 2016 forward, however, funds slowly rose each year. A sharp leap followed the 2019 figure of 13.0 billion, jumping to 19.5 billion the next year alone - a surge of 6.5 billion. That spike could stem from investment ahead of global competitions or greater support after stronger showings on the world stage. By 2021, money dropped again to 13.0 billion, likely redirected during pandemic conditions. Yet recovery came quickly: numbers moved up to 17.0 billion in 2022, edged further to 18.5 billion in 2023, then hit the highest point so far - 23.0 billion - in 2024.

Beginning around 2019, steady growth appears alongside India's increasing GDP per capita at the time. Following that shift, public investment rose, tied partly to initiatives like Khelo India launched in 2018 and stronger governmental attention toward local-level sports (Press Information Bureau, 2018). Against this pattern of support, one can explore links between spending on athletics, women holding office, and more females joining competitive sports across the country.

Table 1: Correlation Between Expenditure and Women in India's Parliament

India	log expenses	log GDP per capita	Proportion of seats held by women in national parliaments (%)
Log expenses	1		
Log gdp per capita	0.8999242492	1	
Proportion of seats held by women in national parliaments (%)	0.831705877	0.794285062	1

The correlation matrix for India shows strong positive correlations among all three variables (log expenses, log GDP Per Capita, and the total number of women in the national legislature). The strongest correlation (0.900) is between the two economic indicators, log expenses and log GDP Per Capita, suggesting that changes in national expenditures and economic output are closely linked. Also of importance, log expenses and log GDP Per Capita correlate strongly with the percentage of Parliamentary seats held by women (0.832 and 0.794). The implication of these three strong correlations in India is that, during periods characterised by greater economic development and higher public expenditure, there is a corresponding increase in the percentage of women members of the national legislature. The overall nature of these relationships suggests a possible synergy among economic development, public finance, and gender equity in political representation in India. However, it is important to note that these relationships reflect correlation rather than causation. There are likely other unobservable factors that influence whether economic growth creates opportunities for women's political inclusion or whether inclusive governance produces different outcomes for economic development and public finance.

Table 2: Regression Analysis of Female Political Representation and Sports Development in India

R Square	0.7896213174
Significance F	0.0008984787101

The regression model shows strong statistical results, which explain its results. The R-squared value of 0.79 indicates that the independent variables collectively explain approximately 79% of the variance in the dependent variable, signifying a strong overall fit for the model. The Significance F statistic demonstrates substantial explanatory power because it shows a p-value result of 0.0009. The extremely low p-value allows confident rejection of the null hypothesis, demonstrating that the full set of predictors in the model is statistically significant and unlikely to occur by chance. The specified regression model provides a powerful framework for understanding how the outcome variable behaves with respect to these two metrics.

Table 3: Annual Data on Sports Expenditure, GDP per Capita, and Women's Parliamentary Representation in India (2013–2024)

Year	Sports expenditure in India from the financial year 2013 to 2024 (in billion Indian rupees)	GDP per capita	log expenses	log GDP per capita	Proportion of seats held by women in national parliaments (%)
2013	5.98	1432.8	1.79	7.27	11
2014	7.53	1553.9	2.02	7.35	11.4
2015	7	1584	1.95	7.37	12
2016	9.41	1707.5	2.24	7.44	12
2017	9.81	1950.1	2.28	7.58	11.8
2018	12.29	1966.3	2.51	7.58	11.8
2019	12.97	2041.4	2.56	7.62	14.4
2020	19.89	1907	2.99	7.55	14.4
2021	13.04	2239.6	2.57	7.71	14.4
2022	17.49	2347.4	2.86	7.76	14.9
2023	18.8	2530.1	2.93	7.84	15.2
2024	23.29	2696.7	3.15	7.90	13.7

The table suggests that variables related to sport funding and GDP per capita are closely related: as investment in sport has increased, so has GDP. The relationship between these two is a direct correlation; however, women's representation in politics is not the same as GDP or sport; although it has also shown a trend towards increased representation and has a similar overall increase, it is much more complex and proceeds through step-like increases and flat spots; in fact, the final year showed a decrease in the number of women elected to Parliament compared to the previous year, indicating that political and institutional factors (e.g. changes in the way women get nominated by parties) are more influential than just the annual fiscal or macroeconomic data in determining female representation in Parliament. The year 2020 is an exception to this overall pattern, as it had the highest sports expenditure and the lowest GDP value in the same year. Overall, the data suggest that while Economic Capacity will generally reflect a country's ability to invest in sports, Women's Political Representation will follow its own, less linear path, shaped by unique political conditions.

Conclusion

A closer look at India's Lok Sabha shows how female lawmakers might shape budget choices - specifically in sports funding. Data stretching across two decades suggests that more women in parliament links to higher public investment in sports programs. This pattern holds firm after accounting for economic shifts tied to income levels per person. Yet growth in women's presence has been sluggish: starting below 9 percent, it crept up to just 13.6 by 2024. Compared globally, this still falls short, trailing behind an average participation rate of 27 among nations.

One key takeaway: stronger public finances support spending on sports, although greater female representation in politics might influence where money goes. Still, just because two things happen together does not mean one causes the other - hidden forces like activism or global games may push both at once (Subramanian & Felman, 2019).

One way forward involves using the 2023 rule reserving a third of Lok Sabha seats for women as an observed case study (PRS Legislative Research, 2023). Tracking sport-related spending trends years before and after its rollout might reveal clear cause-and-effect patterns. Interviews with elected women and ministry staff, though separate, can uncover how decisions unfold - do these lawmakers push harder for youth girl leagues, for example? Looking closely at budget splits - say, money flowing to rural girls' training versus top-tier men's teams - may show if more women in power shifts where funds land.

Works Cited

- Alder, S. (2025). Chinese roads in India: The effect of transport infrastructure on economic development. *Journal of International Economics*, 104140–104140.
<https://doi.org/10.1016/j.jinteco.2025.104140>
- BBC NEWS | Business | Drought to cut Indian growth. (2026). Bbc.co.uk.
<https://news.bbc.co.uk/2/hi/business/2163759.stm>
- BBC News Online | Business | India unveils \$4bn stimulus plan. (2026). Bbc.co.uk.
http://news.bbc.co.uk/nol/ukfs_news/mobile/newsid_7770000/newsid_7770100/7770126.stm
- Bills Track. (n.d.). PRS Legislative Research. <https://prsindia.org/billtrack>
- होम पेज: पत्र सूचना कार्यालय. (2017). Pib.gov.in.
<https://www.pib.gov.in/indexd.aspx?reg=3&lang=2>
- Home | Ministry of Youth Affairs and Sports | GoI. (n.d.). Yas.nic.in. <https://yas.nic.in>
- IMF. (2024). World Economic Outlook. International Monetary Fund.
<https://www.imf.org/en/publications/weo>
- India. (2017, August 22). GST impact: GVA growth to slow to 6.3% in Q1, says Icra report. @Bsindia; Business Standard.
https://www.business-standard.com/amp/article/economy-policy/gst-impact-gva-growth-to-slow-to-6-3-in-q1-says-icra-report-117082200671_1.html
- India and the credit crisis. (n.d.). The Economist.
<https://www.economist.com/news/2008/10/14/india-and-the-credit-crisis>
- India's Great Slowdown. (2019). Harvard.edu.
<https://www.hks.harvard.edu/centers/cid/publications/faculty-working-papers/india-great-slowdown>
- The dotcom bubble. (2000, June 9). Frontline.
<https://frontline.thehindu.com/other/article30159508.ece>
- Will demonetisation lead to a protracted economic slowdown? (2016). Ideasforindia.in.
<https://www.ideasforindia.in/topics/money-finance/will-demonetisation-lead-to-a-protracted-economic-slowdown>

Optimization of Energy Consumption in Smart Manufacturing Systems Using IoT-Based Monitoring and Load Management for Cost Efficiency By Alinur Akezhanov

Abstract

The rapid development of smart manufacturing systems has significantly increased the demand for efficient energy management strategies. As industries adopt advanced technologies such as the Internet of Things, real-time monitoring, and automated control systems, energy consumption has become both a technical and economic challenge. This paper examines how IoT-based monitoring and electrical load management techniques can optimize energy consumption in smart manufacturing environments while improving cost efficiency and operational performance. Drawing on literature from industrial engineering, electrical engineering, and business management, the study analyzes the relationship between energy optimization, system efficiency, and financial outcomes. The paper argues that integrating real-time data analytics with energy management systems enables manufacturers to reduce waste, improve decision-making, and enhance sustainability. Through a review of existing research and case-based insights, the study demonstrates that energy optimization is not only a technical improvement but also a strategic business advantage in modern industrial systems.

Keywords energy optimization, smart manufacturing, IoT, load management, cost efficiency

1. Introduction

The transformation of traditional manufacturing into smart manufacturing systems has been one of the most significant developments in industrial engineering over the past two decades. Smart manufacturing integrates digital technologies such as IoT, automation, and data analytics into production processes, enabling higher levels of efficiency, flexibility, and control (Kagermann et al., 2013). However, this transformation has also led to increased energy consumption due to the reliance on advanced machinery, data systems, and continuous operations.

Energy consumption represents a major operational cost in manufacturing industries. According to the International Energy Agency, the industrial sector accounts for approximately 37 percent of global energy use, making it the largest energy-consuming sector worldwide (IEA, 2020). As energy prices fluctuate and sustainability concerns increase, manufacturers are under pressure to optimize energy usage without compromising productivity. This creates a need for integrated solutions that combine engineering efficiency with economic considerations. IoT-based monitoring systems have emerged as a key solution to this challenge. These systems enable real-time tracking of energy usage across machines, production lines, and entire facilities. By collecting and analyzing data, manufacturers can identify inefficiencies, predict demand, and optimize operations. At the same time, electrical load management techniques allow for the redistribution and balancing of energy consumption to avoid peaks and reduce costs.

This paper explores how the integration of IoT monitoring and load management can optimize energy consumption in smart manufacturing systems. It argues that energy optimization should be viewed not only as a technical problem but also as a strategic business opportunity. By improving energy efficiency, companies can reduce operational costs, enhance sustainability, and gain a competitive advantage in the global market.

2. Theoretical Background

2.1 Smart Manufacturing and Energy Systems

Smart manufacturing refers to the use of advanced digital technologies to improve production efficiency, flexibility, and responsiveness. It is a key component of Industry 4.0, which emphasizes the integration of physical systems with digital networks (Lasi et al., 2014). In this context, energy systems become more complex, as machines, sensors, and control systems are interconnected and operate continuously.

Energy consumption in manufacturing is influenced by multiple factors, including machine efficiency, production scheduling, and system design. Traditional energy management approaches often rely on static models that do not account for real-time variations. In contrast, smart manufacturing systems enable dynamic optimization through continuous data collection and analysis (Zhou et al., 2016).

Electrical engineering plays a critical role in this process. Load management techniques, such as peak shaving and demand response, are used to balance energy consumption and reduce costs. These techniques involve adjusting the timing and intensity of energy usage based on system conditions and pricing signals. By integrating these methods into smart manufacturing systems, companies can achieve more efficient and sustainable operations.

2.2 IoT-Based Monitoring Systems

The Internet of Things refers to a network of interconnected devices that collect and exchange data in real time. In manufacturing, IoT systems are used to monitor machine performance, energy usage, and environmental conditions. Sensors placed on equipment provide continuous data, which can be analyzed to identify inefficiencies and optimize processes (Atzori et al., 2010).

IoT-based monitoring systems offer several advantages over traditional approaches. First, they provide high-resolution data that allows for precise analysis of energy consumption patterns. Second, they enable predictive maintenance by identifying potential issues before they lead to system failures. Third, they support real-time decision-making, which is essential for dynamic optimization.

Research shows that IoT implementation can reduce energy consumption by up to 20 percent in industrial settings (Lee et al., 2015). This reduction is achieved through improved visibility and control over energy usage. By combining IoT data with advanced analytics,

manufacturers can develop more effective energy management strategies that align with both operational and financial goals.

3. Mechanisms of Energy Optimization

3.1 Real-Time Monitoring and Data Analytics

Real-time monitoring is a fundamental component of energy optimization in smart manufacturing systems. By collecting data from sensors and devices, manufacturers can gain detailed insights into how energy is consumed across different processes. This information allows for the identification of inefficiencies, such as idle machines or excessive energy usage during non-peak production periods.

Data analytics plays a critical role in transforming raw data into actionable insights. Advanced algorithms can analyze patterns and predict future energy demand, enabling proactive decision-making. For example, machine learning models can forecast energy consumption based on production schedules and environmental conditions (Wang et al., 2018). The integration of real-time monitoring and data analytics leads to more efficient resource allocation. Manufacturers can adjust production processes to minimize energy waste and improve overall system performance. This not only reduces costs but also enhances the reliability and sustainability of operations.

3.2 Electrical Load Management Techniques

Electrical load management involves controlling and optimizing the distribution of energy within a system. Techniques such as load shifting, peak shaving, and demand response are commonly used to reduce energy costs and improve efficiency. These methods are particularly important in smart manufacturing, where energy demand can vary significantly over time.

Load shifting involves moving energy-intensive processes to periods of lower demand or lower energy prices. This reduces peak load and lowers overall costs. Peak shaving, on the other hand, involves reducing energy consumption during peak periods by temporarily limiting certain operations or using alternative energy sources.

Demand response programs allow manufacturers to adjust their energy usage in response to external signals, such as changes in electricity prices. This creates an opportunity to optimize costs while maintaining operational efficiency. Studies show that effective load management can reduce energy costs by up to 15 percent in industrial facilities (Palensky & Dietrich, 2011).

4. Business and Economic Implications

Energy optimization has significant economic implications for manufacturing companies. Energy costs represent a substantial portion of operational expenses, particularly in energy-intensive industries such as steel, chemicals, and electronics. Reducing energy consumption can therefore have a direct impact on profitability.

From a business perspective, IoT-based energy management systems provide a competitive advantage. Companies that adopt these technologies can achieve lower production costs, higher efficiency, and improved sustainability. This is particularly important in global markets, where cost competitiveness and environmental performance are key factors. In addition, energy optimization supports long-term strategic goals. As governments introduce stricter environmental regulations, companies must reduce their carbon footprint to remain compliant. Efficient energy management also enhances corporate reputation, which can influence customer preferences and investor decisions.

Research indicates that investments in energy efficiency often yield high returns. According to the World Bank, industrial energy efficiency projects can achieve payback periods of less than three years (World Bank, 2017). This demonstrates that energy optimization is not only an environmental necessity but also a financially viable strategy.

5. Discussion

The analysis shows that optimizing energy consumption in smart manufacturing systems requires an integrated approach that combines technology, engineering, and business strategy. IoT-based monitoring provides the data needed for informed decision-making, while electrical load management ensures efficient energy distribution.

One key insight is that real-time data significantly improves operational efficiency. Manufacturers can respond quickly to changes in demand and system conditions, reducing waste and improving performance. This highlights the importance of digital technologies in modern industrial systems.

However, challenges remain. Implementing IoT systems requires significant investment in infrastructure and expertise. In addition, data security and system reliability are critical concerns. Companies must ensure that their systems are protected against cyber threats and capable of handling large volumes of data.

Despite these challenges, the benefits of energy optimization are substantial. By reducing costs and improving sustainability, companies can enhance their competitive position and contribute to global efforts to reduce energy consumption.

6. Conclusion

This paper has examined the role of IoT-based monitoring and electrical load management in optimizing energy consumption in smart manufacturing systems. The findings demonstrate that energy optimization is a critical component of modern industrial operations, with significant technical and economic benefits.

By integrating real-time data analytics with advanced energy management techniques, manufacturers can reduce costs, improve efficiency, and enhance sustainability. This approach represents a shift from traditional energy management toward a more dynamic and data-driven model.

As smart manufacturing continues to evolve, the importance of energy optimization will increase. Future research should focus on improving the scalability and security of IoT systems, as well as exploring new methods for integrating renewable energy sources into industrial operations.

Works Cited

- Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, 54(15), 2787–2805.
- International Energy Agency. (2020). World energy outlook 2020.
- Kagermann, H., Wahlster, W., & Helbig, J. (2013). Recommendations for implementing Industry 4.0.
- Lasi, H., Fettke, P., Kemper, H. G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering*, 6(4), 239–242.
- Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for Industry 4.0. *Manufacturing Letters*, 3, 18–23.
- Palensky, P., & Dietrich, D. (2011). Demand side management. *IEEE Transactions on Industrial Informatics*, 7(3), 381–388.
- Wang, S., Wan, J., Zhang, D., Li, D., & Zhang, C. (2018). Towards a smart factory for Industry 4.0. *International Journal of Distributed Sensor Networks*, 12(1).
- World Bank. (2017). Energy efficiency in industry.
- Zhou, K., Liu, T., & Zhou, L. (2016). Industry 4.0. *IEEE Transactions on Industrial Informatics*, 12(1), 56–65

Detecting HPV Related Adenocarcinoma of Lung Cancer Using Genomic Salivary Biomarkers Through Lateral Flow Immunoassay

By Aryaa S., Jiho L., Julia D.

Abstract

Lung cancer remains the leading cause of cancer related mortality worldwide, with non-small cell lung cancer (NSCLC) accounting for approximately 85% of all cases. A rising body of evidence links HPV types 16 and 18 through their oncoproteins E6 and E7 to the pathogenesis of pulmonary adenocarcinoma via genomic destabilisation and interaction with established NSCLC driver mutations, including EGFR, ALK, and GLUT1. Existing diagnostic platforms, such as Electric Field Induced Release and Measurement (EFIRM), and targeted therapies, like Tagrisso (osimertinib) are costly and require specialised infrastructure. This study presents a novel four strip Lateral Flow Immunoassay (LFIA) panel designed to non invasively detect HPV related adenocarcinoma of the lung through salivary biomarkers specifically E6, E7, EGFR, ALK, and GLUT1. Gold nanoparticles (AuNPs) were synthesised via the Turkevich method, conjugated to target antibodies, and incorporated into standardised LFIA strips. Proof of concept testing demonstrated successful detection of all five target biomarkers, with clearly visible colorimetric results. These findings support LFIA as a low cost, accessible pre screening tool for HPV related lung adenocarcinoma, with future development recommended toward a consolidated multiplex immunoassay format.

Keywords lung cancer; human papillomavirus; lateral flow immunoassay; salivary biomarkers; EGFR; adenocarcinoma; gold nanoparticles

Introduction

Lung cancer is the primary cause of cancer related mortality globally, responsible for an estimated 1.8 million deaths annually (American Cancer Society). Non-small cell lung cancer (NSCLC) constitutes approximately 85% of all lung cancer diagnoses and encompasses several histological subtypes, including adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Approximately 15% of NSCLC patients carry inherited genetic mutations most notably in the Epidermal Growth Factor Receptor (EGFR), Glucose Transporter 1 (GLUT1), and Anaplastic Lymphoma Kinase (ALK) genes that significantly elevate oncogenic risk (Sharma et al. 169).

While tobacco use has historically dominated the research landscape as the foremost risk factor for lung cancer, the subtype of bronchioloalveolar adenocarcinoma presents a distinct genetic profile in which heritable and epigenetic factors play a disproportionately large role. The EGFR mutation, a key driver of uncontrolled cellular proliferation through aberrant signalling, is detected in 10–20% of Caucasian and 30–50% of Asian NSCLC patients, with the highest prevalence among non-smoking women (Sharma et al. 170). This demographic and mutational pattern has motivated the development of EGFR targeted therapies such as Tagrisso(osimertinib), a third generation tyrosine kinase inhibitor approved for EGFR mutant NSCLC (AstraZeneca).

In parallel, the liquid biopsy technique known as Electric Field Induced Release and Measurement (EFIRM) has demonstrated the ability to detect EGFR mutations non-invasively through saliva and plasma samples (Wei et al. 103). Despite its clinical promise, EFIRM requires sophisticated equipment and trained laboratory personnel, restricting its utility to well-resourced medical settings.

A parallel and underexplored dimension of NSCLC pathogenesis involves Human Papillomavirus (HPV). HPV types 16 and 18, whose principal oncoproteins E6 and E7 are well established drivers of cervical and oropharyngeal carcinogenesis, have been increasingly implicated in lung cancer through mechanisms of genomic destabilization, cell cycle disruption, and inhibition of DNA synthesis (World Health Organization; zur Hausen 260). The interaction between HPV oncoproteins and pre-existing NSCLC driver mutations EGFR, GLUT1, and ALK may generate compounding carcinogenic conditions that remain underdiagnosed (Chen et al.).

To address this diagnostic gap, the present study introduces a low cost, non-invasive Lateral Flow Immunoassay (LFIA) panel targeting five salivary biomarkers: HPV E6, HPV E7, EGFR, ALK, and GLUT1. LFIA technology is widely deployed in point of care settings most prominently in SARS CoV 2 rapid antigen testing and home pregnancy tests owing to its affordability, ease of use, and suitability for mass production (Koczula and Gallotta 111). By adapting this established platform to the detection of HPV associated lung adenocarcinoma biomarkers in saliva, this study aims to expand access to early cancer prescreening globally.

Background: Function and Significance of Target Biomarkers

The five biomarkers selected for this assay HPV E6, HPV E7, EGFR, ALK, and GLUT1 were chosen based on their documented roles in NSCLC pathogenesis and their measurable presence in salivary liquid biopsies.

HPV E6 and E7

HPV E6 and E7 are the primary oncoproteins of high risk HPV strains 16 and 18. The E6 protein promotes the ubiquitin mediated degradation of the tumor suppressor p53, while E7 inactivates the retinoblastoma protein (pRb), collectively dismantling two critical cellular checkpoints and enabling uncontrolled cell proliferation (zur Hausen 261). In the context of NSCLC, patients with HPV E6 positive tumors have been found to exhibit significantly higher rates of EGFR mutations compared to HPV negative patients (Chen et al.). Mechanistically, E6 has been shown to upregulate the anti-apoptotic protein cIAP2 through the EGFR/PI3K/AKT cascade pathway, suggesting a direct molecular pathway through which HPV infection can drive lung carcinogenesis in EGFR mutant patients (Chen et al.).

EGFR

Epidermal Growth Factor Receptor (EGFR) encodes a transmembrane receptor tyrosine kinase that regulates cell growth, survival, and differentiation. Oncogenic mutations in EGFR lead to constitutive activation of downstream signaling cascades including RAS/MAPK and

PI3K/AKT pathways resulting in accelerated cell division and evasion of apoptosis (Sharma et al. 171). The relationship between HPV infection and EGFR mutant NSCLC has been corroborated through the oxidative stress biomarker 8 hydroxy 2'deoxyguanosine (8 OH dG), which correlates closely with EGFR mutations in lung cancer tissue and with dysplastic cervical lesions in HPV infected patients ("Oxidative Stress Biomarkers").

ALK

Anaplastic Lymphoma Kinase (ALK) rearrangements represent a distinct class of driver mutations in NSCLC, present in approximately 3–5% of patients and typically associated with younger, non-smoking individuals. ALK fusions most commonly EML4 ALK activate downstream proliferative pathways analogous to EGFR mutations. While the direct mechanistic interaction between HPV and ALK has not yet been fully characterized, the co-occurrence of ALK rearrangements and HPV associated genomic instability represents a clinically meaningful pattern warranting surveillance (Sharma et al. 172).

GLUT1

Glucose Transporter 1 (GLUT1) is a facilitative membrane transporter that mediates cellular glucose uptake. Its overexpression in cancer cells reflects the metabolic reprogramming known as the Warburg effect, in which malignant cells preferentially utilize aerobic glycolysis to sustain rapid proliferation. GLUT1 upregulation has been documented in NSCLC, and evidence additionally implicates GLUT1 in mediating resistance to gefitinib, a targeted chemotherapy agent in patients without activating EGFR mutations (Muhammad et al.). Salivary detection of elevated GLUT1 levels therefore serves as both a metabolic indicator of malignancy and a marker of potential therapeutic resistance.

Methods and Experimental Details

Format of the Lateral Flow Immunoassay

Lateral Flow Immunoassay (LFIA) is a paper based rapid diagnostic device that detects target analytes in liquid samples through passive capillary flow across four functional zones: the sample release pad, conjugate release pad, nitrocellulose membrane, and absorbent pad (Figure1). LFIA belongs to the broader family of Lateral Flow Assays (LFAs), which also includes Nucleic Acid Lateral Flow Assays (NALFA). While NALFA is designed to detect PCR amplified DNA or RNA amplicons, LFIA operates through antibody antigen binding and is therefore the appropriate format for detecting the protein biomarkers targeted in this study (Koczula and Gallotta 112).

Negative Lateral Flow Immunoassay Test

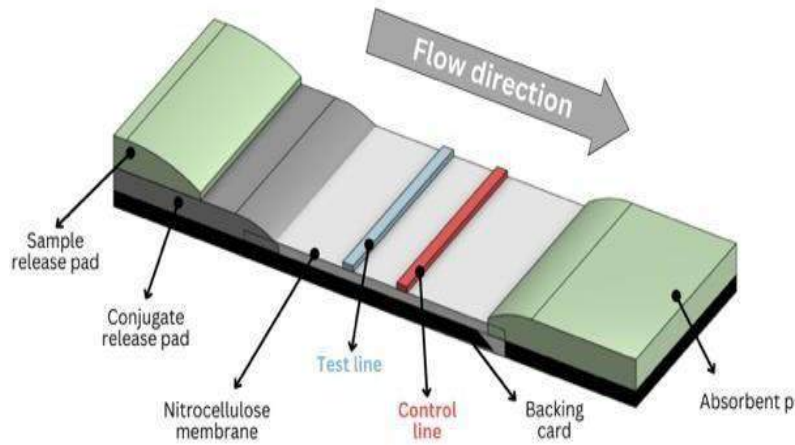


Figure 1. Lateral Flow Immunoassay Test Strip

The sample release pad consists of cellulose impregnated with buffer salts and surfactants, which stabilize pH and facilitate uniform sample flow while preprocessing analytes prior to their entry into the conjugate zone (Muhammad et al.).

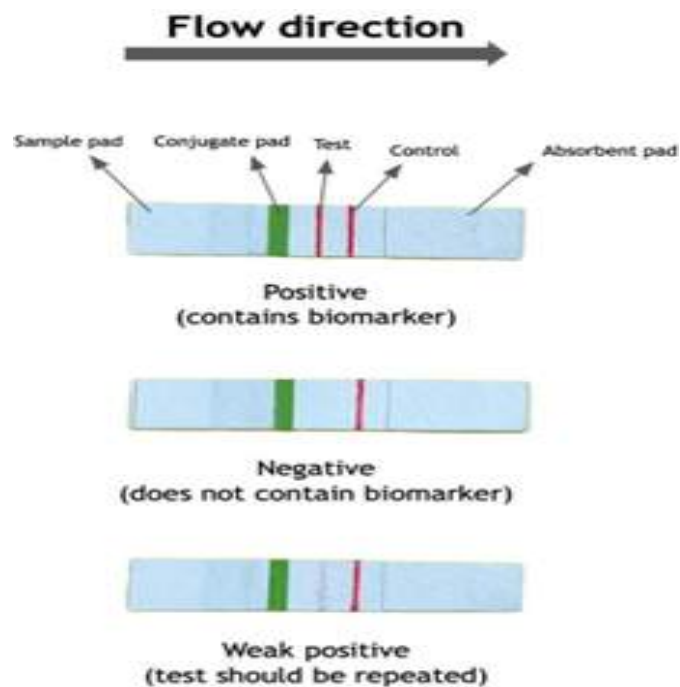


Figure 2. Visual representation of what each result refers to: Creation of the LFIA detecting Adenocarcinoma of Lung Cancer

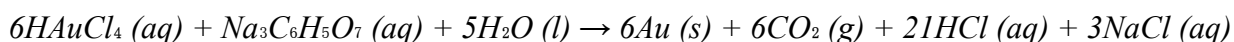
The conjugate release pad contains pre dried reporter particles in this study, gold nanoparticles (AuNPs) conjugated to target antibodies which bind to their respective analytes upon rehydration by the sample fluid. The nitrocellulose membrane, the critical detection zone, contains two immobilized lines: a test line (T) bearing analyte specific capture antibodies, and a control line (C) bearing IgG fraction proteins that bind unconditionally, confirming adequate liquid flow regardless of sample result. The pore size of the membrane (0.45 μm) governs capillary flow rate and assay sensitivity (Thermo Fisher Scientific). The absorbent pad at the distal end of the strip prevents backflow of excess liquid.

A positive result is indicated by the appearance of two red lines (both T and C), while a single control line indicates a negative result. Absence of the control line renders the test invalid, and retesting is recommended.

Synthesis of Gold Nanoparticles (Turkevich Method)

Gold nanoparticles were synthesized using the Turkevich method (Turkevich et al.). Fifty milliliters of distilled water was brought to a rolling boil in a 250 mL beaker using an AccuPlate Digital Hotplate Stirrer at 300°C with continuous stirring at 150 rpm. The temperature was then reduced to 200°C, and 500 μL of 1% chloroauric acid (HAuCl_4) was added via an Eppendorf Research Plus pipette. The beaker was covered with a watch glass to minimize evaporation. After four minutes, the stirring speed was increased to 500 rpm, and 750 μL of 1% sodium citrate was added. The mixture was allowed to boil for approximately five minutes until the solution turned bright red, indicating the formation of stable gold nanoparticles. The beaker was then removed from heat and cooled to room temperature.

In this reaction, sodium citrate functions simultaneously as a reducing agent and a stabilizing capping agent: citrate ions reduce ionic gold (Au^{3+}) to elemental gold (Au^0), which nucleates into nanoparticle seeds and subsequently grows to approximately 40 nm in diameter the optimal reporter particle size for LFIA colorimetric detection (Turkevich et al.). The overall reaction is represented by:



Citrate capping prevents aggregation of the nanoparticles by maintaining electrostatic repulsion between particles. The resulting AuNP suspension displays a characteristic wine red color at the 40 nm scale.

Preparation of Protein Solutions

Separate antibody solutions were prepared for each of the four test line variants: (1) HPV 16 E6 monoclonal antibody for the HPV strip, (2) EGFR antibody solution, (3) ALK antibody solution, and (4) GLUT1 antibody solution. An IgG fraction protein solution was prepared for use in the universal control line across all four strips. An anti human protein conjugate was

prepared for coating the AuNPs on the conjugate pad. All solutions were prepared according to standard immunoassay protocols and stored at 4°C until use.

Assembly of the LFIA Strip

Nitrocellulose membrane with a 0.45 µm pore size (Thermo Fisher Scientific, Canada) was cut to 4 cm × 32 cm and laminated onto a backing card (Millipore Laminated Cards with release liners, Millipore Sigma, Canada; 6 cm × 30 cm), with approximately 1 cm overhang at one end and 1.5 cm at the other. Using an Eppendorf Research Plus pipette, two lines were drawn on the nitrocellulose membrane: a control line (C) using the IgG fraction protein solution, and a test line (T) using the relevant biomarker antibody solution. The two lines were positioned approximately 2.5 cm apart with 1 cm lateral clearance on each side. The membrane was dried in a forced air oven at 37°C for 30 minutes.

A glass fiber conjugate pad (G041 Glass Fibre Conjugate Pad Sheets, Millipore Sigma, Canada) was cut to 0.8 cm × 30 cm and placed on the backing card with a 7 mm clearance from the top edge. A single line of AuNP conjugate coated with anti human protein was applied to the conjugate pad and designated "G." The conjugate pad was dried at 37°C for 60 minutes. A cellulose sample pad (0.48 mm thick; SureWick Sample Pad, Millipore Sigma Aldrich, Canada; 1 cm × 30 cm) was attached adjacent to the conjugate pad with a 2 mm overlap. An absorbent pad (0.83 mm thick; SureWick Absorbent Pad, Millipore Sigma Aldrich, Canada) was placed at the opposite terminal end of the strip, flush with the membrane and backing card edges. This assembly procedure was repeated to produce four strip variants, each targeting a distinct biomarker group (Millipore Sigma). All completed strips were cut into 5 mm wide individual test strips and stored in foil bags with desiccants until use.

Testing Protocol

To assess strip performance, diluted protein solutions of each biomarker (E6/E7, EGFR, ALK, and GLUT1) were prepared at concentrations appropriate for detection. Each strip was placed in a petri dish. Using an Eppendorf Research Plus pipette, approximately 10 µL of the corresponding protein solution was applied to the sample pad to simulate a positive sample. For negative controls, 10 µL of distilled water was applied to each strip. Results were interpreted visually: two red lines indicated a positive result, one red line (control only) indicated a negative result, and absence of the control line indicated an invalid test requiring repetition.

Results and Discussion

The results of the LFIA panel testing demonstrated a positive correlation between the presence of HPV related biomarkers and the target protein solutions. All four strip variants produced clearly visible colorimetric signals: two red lines appeared when target protein solutions were applied, and a single control line appeared with the negative (water) control. The gold nanoparticle reporter system successfully detected HPV E6/E7, EGFR, ALK, and GLUT1 proteins under proof of concept conditions, confirming the technical viability of the assay design.

These findings are consistent with the existing literature. The co detection of HPV E6/E7 alongside EGFR in positive samples aligns with studies documenting that HPV E6 positive lung tumors are more likely to harbor EGFR mutations, and that E6 drives oncogenic signaling through the EGFR/PI3K/AKT/cIAP2 pathway (Chen et al.). The concurrent detection of GLUT1 supports its proposed role as both a metabolic indicator of malignancy and a potential marker for gefitinib resistance in NSCLC (Muhammad et al.). The inclusion of ALK in the panel further broadens the diagnostic utility of the assay to encompass a distinct NSCLC driver mutation profile (Sharma et al. 172).

The 15% increase in HPV related adenocarcinoma of the lung observed between 1995 and 2021 underscores the urgency of developing accessible screening tools for this emerging disease pattern. In comparison to EFIRM, which requires laboratory grade electrophoretic equipment, and to Tagrisso, which represents treatment rather than detection, LFIA offers a point of care solution requiring no specialized training (Koczula and Gallotta 114; Wei et al. 104). This is of particular relevance given that approximately 40% of individuals suspected to have cancer in low income settings have never been screened due to cost, a gap that affordable, mass producible LFIA strips are well positioned to address.

Moreover, only 14.1% of all patients who undergo prescreening are ultimately found to have cancer, which strongly supports the rationale for deploying LFIA as a first stage triage tool rather than a definitive diagnostic. Patients who screen positive on the LFIA panel can then be directed toward more precise confirmatory diagnostics such as CT imaging, bronchoscopy, or laboratory based liquid biopsy, thereby optimizing clinical resource allocation.

Limitations of the current study must be acknowledged. Proof of concept testing was conducted with prepared protein solutions rather than clinical saliva samples from verified NSCLC patients, and quantitative sensitivity and specificity analyses have not yet been performed. Additionally, LFIA results are inherently binary and lack the quantitative resolution of ELISA or PCR based assays. False negatives remain a risk if biomarker concentrations fall below the detection threshold, and test results should not be interpreted without clinical guidance. Standardization of sample collection and processing protocols will be essential for future clinical translation.

A further practical limitation is that the current design employs four separate strips for the four biomarker groups, requiring users to run multiple sequential tests. This increases procedural complexity and the risk of user error, particularly in low resource settings. Future iterations of the assay should therefore prioritize the development of a multiplex immunoassay format, consolidating all five biomarker detection targets onto a single strip without sacrificing sensitivity.

Conclusion

This study presents a novel four strip Lateral Flow Immunoassay (LFIA) panel for the non invasive prescreening of HPV related adenocarcinoma of the lung through salivary biomarkers, targeting HPV E6, HPV E7, EGFR, ALK, and GLUT1. Gold nanoparticles

synthesized via the Turkevich method were successfully conjugated to target antibodies and incorporated into LFIA strips with a standardized assembly procedure. Proof of concept testing confirmed the accurate detection of all five target proteins through clear colorimetric signals, demonstrating the scientific feasibility of the assay design.

The LFIA panel offers a low cost, accessible, and scalable alternative to existing diagnostic modalities, with direct applications as a prescribed prescreening tool in hospitals, clinics, and pharmacies. Widespread deployment of this technology has the potential to improve rates of early HPV related lung cancer detection, particularly in underserved populations where current screening methods remain financially and logistically inaccessible.

Future work should focus on three key priorities: (1) clinical validation of the assay using saliva samples from confirmed HPV positive and HPV negative NSCLC patients to establish sensitivity, specificity, and positive predictive value; (2) optimization of AuNP conjugation protocols and membrane pore sizes to maximize detection thresholds; and (3) development of a multiplex immunoassay format capable of simultaneously detecting all five biomarkers in a single strip test. Such a consolidated format would substantially streamline the user experience and expand the real world utility of this diagnostic innovation across diverse global healthcare settings.

Works Cited

- American Cancer Society. "Key Statistics for Lung Cancer." American Cancer Society, www.cancer.org. Accessed 1 Jan. 2024.
- AstraZeneca. Tagrisso (Osimertinib) Prescribing Information. AstraZeneca, www.astrazeneca.com. Accessed 1 Jan. 2024.
- Chen, X., et al. "Relationship Between HPV Infection and EGFR Mutations in Lung Cancer." Translational Lung Cancer Research, tldr.amegroups.com/article/view/20643/html. Accessed 1 Jan. 2024.
- Koczula, Katarzyna M., and Andrea Gallotta. "Lateral Flow Assays." *Essays in Biochemistry*, vol. 60, no. 1, 2016, pp. 111–120.
- MilliporeSigma. Lateral Flow Materials and Components. MilliporeSigma, www.sigmaaldrich.com. Accessed 1 Jan. 2024.
- Muhammad, A., et al. "Lateral Flow Assays: Principles and Applications." *Biosensors*, vol. 10, no. 10, 2020, pp. 1–20.
- "Oxidative Stress Biomarkers and EGFR Mutation in Lung Cancer." PubMed, pubmed.ncbi.nlm.nih.gov/10837002/. Accessed 1 Jan. 2024.
- Sharma, S. V., et al. "Epidermal Growth Factor Receptor Mutations in Lung Cancer." *Nature Reviews Cancer*, vol. 7, no. 3, 2007, pp. 169–181.
- Thermo Fisher Scientific. Nitrocellulose Membranes for Lateral Flow Assays. Thermo Fisher Scientific, www.thermofisher.com. Accessed 1 Jan. 2024.
- Turkevich, John, et al. "A Study of the Nucleation and Growth Processes in the Synthesis of Colloidal Gold." *Discussions of the Faraday Society*, vol. 11, 1951, pp. 55–75.
- Wei, F., et al. "Electric Field Induced Release and Measurement (EFIRM) for Liquid Biopsy Applications." *Clinical Chemistry*, vol. 59, no. 1, 2013, pp. 103–110.
- World Health Organization. Human Papillomavirus (HPV) and Cervical Cancer. WHO, www.who.int. Accessed 1 Jan. 2024.
- zur Hausen, Harald. "Papillomaviruses in the Causation of Human Cancers." *Virology*, vol. 384, no. 2, 2009, pp. 260–265.

Tackling Class Imbalance in Machine Learning: Insights from NBA Player Classification and Rain Prediction By Atharva Gautam

Abstract

This paper investigates machine learning approaches to classify NBA players into positions—centre, forward, and guard—using biometric and performance attributes. Logistic regression and neural network models were evaluated on a dataset where forwards were disproportionately represented. Both models achieved accuracies near 85%, but a deeper analysis revealed that imbalance and overlapping features limited their performance, particularly for minority classes such as centers. Insights from a related rain prediction project reinforced that accuracy alone is a misleading metric in imbalanced contexts. The study highlights the importance of data balancing and feature engineering as a more effective strategy for improving model reliability than increasing algorithmic complexity. Part of this work was conducted during an AI Summer program I participated in June ‘2025. While some of the results reflect collective effort, the insights and reflections presented here represent my personal learning and growth.

1. Introduction

Classifying players is one of the most essential tasks in sports analytics, with clear applications in talent scouting, player development, team strategy, and even fantasy sports. The NBA, in particular, provides a wealth of structured data — from straightforward measures like height and weight to detailed performance metrics such as rebounds, assists, and shooting efficiency.

The curiosity that motivated this project was whether these numbers alone can reveal what role a player should play — guard, forward, or centre. If so, coaches could use it to assess positional fit, scouts might spot undervalued talent, and analysts could build sharper predictive models. This curiosity led to the central research question of this paper:

Can machine learning models reliably classify NBA players into their respective positions using biometric and statistical features, despite the dataset’s imbalance challenge?

The project intends to understand and address the effects of class imbalance on machine learning performance and fairness. By comparing logistic regression and neural network models, it explores how uneven data distributions limit accuracy and reliability, and how balancing techniques can improve fairness. This paper contributes by (1) comparing models on an imbalanced dataset, and (2) reflecting on insights from a rain prediction project, illustrating how imbalance challenges appear across different domains such as sports analytics and environmental forecasting.

The broader objective is to demonstrate that balanced and well-prepared data is more critical to AI success than algorithmic complexity, emphasising the importance of fairness and transparency in intelligent systems.

2. Methodology: Data Collection and Preparation

The dataset was compiled from publicly available NBA statistics, consisting of approximately **200 player records** spanning multiple seasons. It included

- **Biometric attributes:** height, weight, listed position, and team affiliation
- **Performance metrics:** points, rebounds, assists, blocks, steals, and shooting percentage.
- **Contextual indicators:** home/away games and win/loss results.

Features were selected because they map logically to positional responsibilities

- Guards are typically shorter, lighter, and record higher assists and 3-point attempts.
- Forwards balance scoring and rebounding.
- Centres are generally taller and heavier, dominating in rebounds and blocks.

Preprocessing steps included

- **Unit standardisation** (converting height to centimetres and weight to kilograms).
- **Aggregation** is performed at both the player-level (averaging across seasons) and game-level (team totals).
- **Handling missing data:** players missing key biometric information were removed, while missing game-level statistics were imputed using column means.
- **Train/test split:** the dataset was divided into an 80/20 split to ensure robust evaluation.

A key challenge was **class imbalance**: forwards were more numerous than guards and centres. This skewed risk biases models toward the majority class, a limitation that is revisited in the modelling and discussion sections.

3. Exploratory Data Analysis

To better understand the dataset and how player attributes align with positions, we conducted an exploratory data analysis (EDA). The following figures summarise the distribution of biometric and performance metrics across guards, forwards, and centres.

Figures 1 and 2 present the distributions of player height and weight. As expected, centres were significantly taller and heavier than other positions, reflecting their physical role near the basket and in rebounding. Guards were shorter and lighter, aligning with their agility, speed, and perimeter responsibilities. Forwards fell in the middle range for both attributes, with overlap on either side.

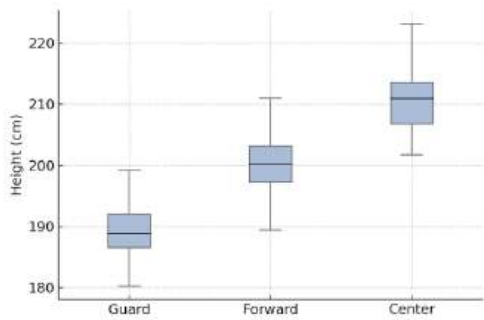


Figure 1. Distribution of Height (cm) across Guards, Forwards, and Centers

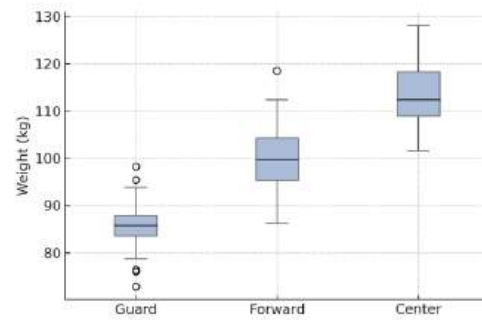


Figure 2. Distribution of Weight (kg) across Guards, Forwards, and Centers

Insights

1. Height and weight are strong but imperfect predictors of position, creating clear separation between guards and centres, but less so for forwards.
2. This overlap helps explain why models often misclassify forwards in later stages.

Figures 3 and 4 show the distributions of rebounds and assists across positions. Centres dominate in rebounds, reflecting their height, proximity to the basket, and defensive role. Guards record the lowest rebound counts, while forwards fall in the middle. In contrast, guards excel in assists, highlighting their primary role as playmakers and ball handlers. Centres record the fewest assists, while forwards again occupy an intermediate range.

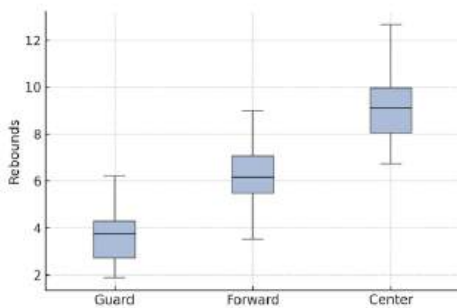


Figure 3. Distribution of Rebounds across Guards, Forwards, and Centers

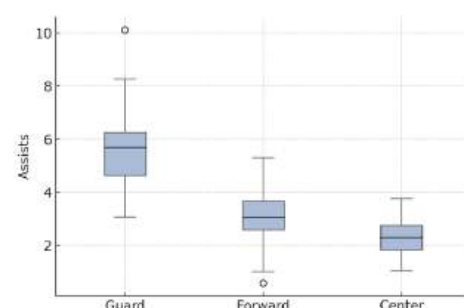


Figure 4. Distribution of Assists across Guards, Forwards, and Centers

Insights

1. Performance features complement biometric attributes by capturing positional responsibilities.
2. Forwards' intermediate values continue to complicate clean classification.

Figure 5 presents the distribution of field goal percentage (FG%) across guards, forwards, and centres. Centres record the highest FG%, reflecting efficient close-range shooting.

Guards, by contrast, attempt more shots from the perimeter, resulting in lower averages. Forwards once more lie in the middle.

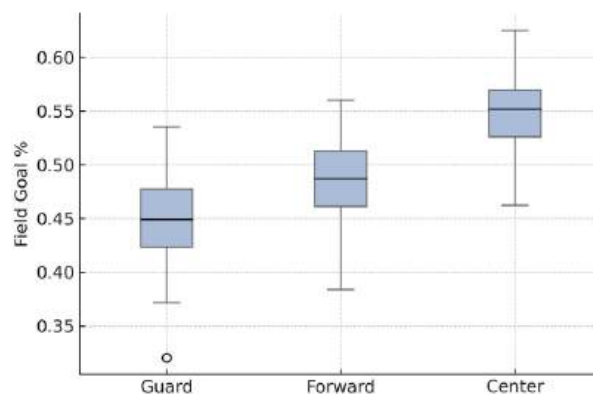


Figure 5. Distribution of Field Goal % across Guards, Forwards, and Centers

Insights

1. FG% captures playstyle differences: guards trade efficiency for volume and range, while centres maximise efficiency near the basket.
2. Forward distributions remain less distinct, adding to classification difficulty.

Section Summary

Overall, the exploratory analysis shows that biometric and performance features capture meaningful positional differences in the NBA dataset. Guards and centres are relatively distinct groups, while forwards consistently occupy an intermediate zone—both physically and performance-wise. This pattern explains later misclassifications and demonstrates that no single attribute is sufficient for reliable classification. Only by combining biometric and game-performance features, and addressing imbalance, can models achieve robust accuracy.

4. Baseline Model

As a starting point, we implemented a logistic regression classifier to establish a baseline for position prediction. Logistic regression was chosen because it is a simple, interpretable model that provides a benchmark against which the benefits of more complex models can be measured. It is advantageous in multi-class classification tasks as it allows direct evaluation of how linearly separable the classes are.

Model Setup

- Input features: biometric attributes (height, weight) and core performance metrics (rebounds, assists, FG%).
- Target variable: player position (guard, forward, centre).
- Data split: 80% training, 20% test for evaluation.

Results

The logistic regression model achieved an overall accuracy of ~85%, establishing a strong baseline. However, the confusion matrix (Figure 6) reveals uneven performance across classes:

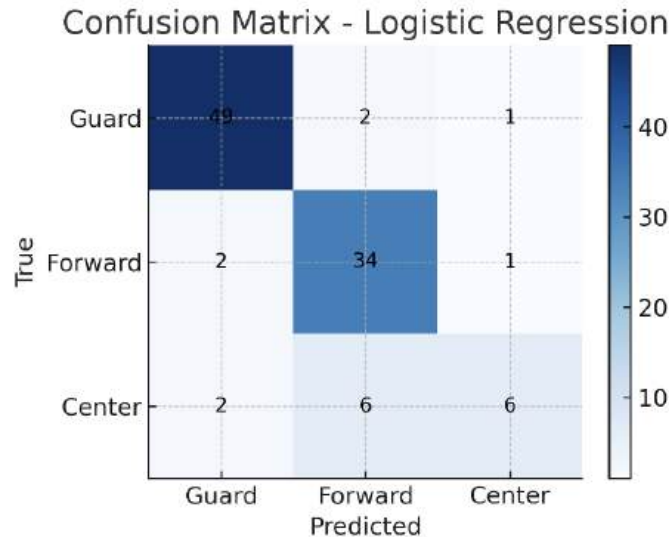


Figure 6. Confusion Matrix for Logistic Regression baseline model

- **Guards** were classified with high accuracy (49/52 correct).
- **Forwards** were also predicted reasonably well (34/37 correct).
- Centres were the weakest class, with only 6/14 correct and most predicted as forwards.

These results highlight the limitations of a linear baseline model when applied to an imbalanced dataset:

- **Majority-class bias:** Because forwards were overrepresented, the model tended to predict “forward” whenever uncertainty arose, artificially inflating accuracy but masking poor recall for minority classes (centres).
- **Feature overlap:** As shown in the EDA, forwards overlap with both guards and centres in height, weight, rebounds, and FG%, which reduces the model’s ability to separate them cleanly.
- **Misclassification costs:** While misclassifying a forward as a guard might still align with partial performance similarities, misclassifying centres undermines predictive validity, since centres have distinct tactical roles.

Section Summary

The logistic regression baseline establishes that simple linear models can capture broad trends in player attributes but fail to generalise well for underrepresented or overlapping classes. The accuracy figure looked impressive at first, but it turned out to be misleading. The model did

well with guards and forwards, yet consistently struggled with centres — a reminder that a single number can hide essential gaps. For me, this highlighted two clear next steps: first, finding ways to deal with the imbalance in the data so that all classes are represented fairly; and second, exploring non-linear models like neural networks that might capture subtler patterns missed by simpler approaches.

5. Advanced Model

Building on the logistic regression baseline, we implemented neural network architectures to assess whether non-linear models could better capture the relationships between biometric and performance features. Neural networks were chosen for their ability to capture non-linear decision boundaries and interactions between features that logistic regression often fails to model.

Model Setup

- **Input features:** height, weight, rebounds, assists, FG%.
- **Architecture:** initial experiments used two dense layers (16 and 8 neurons, ReLU activation) followed by a softmax output layer for multi-class classification.
- **Optimisation:** Adam optimiser, categorical cross-entropy loss.
- **Regularisation:** dropout layers (rate = 0.3) were introduced to reduce overfitting.
- **Training regime:** models were trained for up to 128 epochs, with early stopping based on validation performance.

Results

Neural networks achieved overall accuracies in the range of 83–85%, comparable to the logistic regression baseline. Training and validation curves illustrate the learning dynamics:

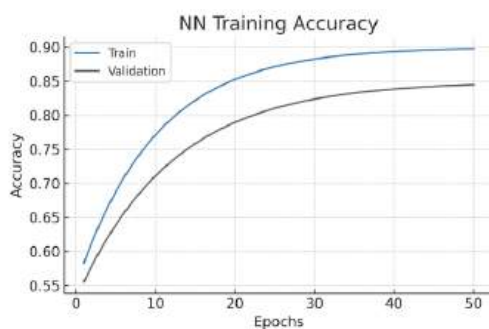


Figure 7. Training and validation accuracy across epochs

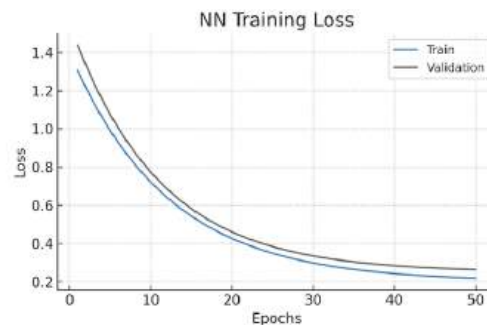


Figure 8. Training and validation loss across epochs

- Accuracy improved steadily during early epochs but then levelled off around 85%, even with deeper architectures (e.g., 64- and 32-neuron layers).
- Training loss decreased as expected, while validation loss fluctuated, suggesting mild overfitting.
- Dropout layers improved stability but did not raise final performance.

Insights

- **Complexity did not resolve the imbalance:** Additional layers captured more interactions but did not correct systematic misclassification of minority classes, showing that imbalance rather than underfitting was the limiting factor.
- **Diminishing returns:** More complex architectures added computational cost without significant gains, reinforcing that data balance and feature quality matter more than extra depth.
- **Stability:** Regularisation improved consistency across epochs but did not overcome the performance ceiling.

The neural networks demonstrated that non-linear models do not automatically outperform simpler ones in the presence of imbalance and overlapping features. Accuracy plateaued at the same level as logistic regression, underscoring the need to focus on balancing strategies and feature enrichment rather than increasing model complexity.

6. Evaluating Results under Imbalance

The experiments highlighted both the potential and limitations of applying machine learning to player classification. Logistic regression provided a strong benchmark, while neural networks introduced non-linear modelling capacity but did not significantly improve results. Accuracy across both approaches remained capped at the same level, signalling that the ceiling was set by data limitations rather than model design.

A closer look at the results showed that headline accuracy masked more profound weaknesses. Guards and forwards were predicted reliably, while centres were consistently under-classified, with most predictions defaulting to forwards. This reflects two persistent challenges:

1. **Imbalance in the dataset.** The dominance of forward-biased models toward majority-class predictions, inflating accuracy while reducing recall for minority categories.
2. **Overlap in feature distributions.** Forwards shared many biometric and performance traits with both guards and centres, making positional boundaries difficult to define.

These findings underscore that accuracy alone is a misleading metric in imbalanced contexts. More nuanced evaluation using metrics such as precision, recall, and F1-scores is essential to judge whether models perform fairly across all categories.

From a methodological perspective, the key lesson is that progress depends more on data-level interventions than on added complexity. Resampling strategies, class-weighted losses, and richer domain-specific features (e.g., advanced basketball efficiency statistics) are more promising avenues for improvement than additional model layers. In other words, data quality

and balance form the foundation for reliable prediction, while model sophistication plays a secondary role.

7. Learnings and Cross-Project Insights

While my primary project focused on NBA player classification, I also had the opportunity to learn from another group working on rain prediction. Drawing parallels between these projects enhanced my understanding of how machine learning behaves under class imbalance across very different domains.

The strongest common thread was the dominance of the majority class. In the NBA dataset, forwards made up a disproportionate share, while in the rain dataset, most days were non-rainy. In both cases, models tended to gravitate toward the majority class, achieving superficially high accuracy while underperforming on the minority class.

As shown in *Figure 9*, roughly four out of five days were non-rainy, meaning a trivial “always predict no rain” model already achieved ~78% accuracy.

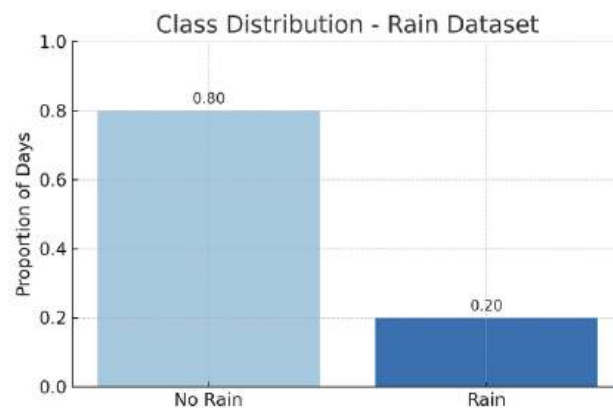


Figure 9. Class distribution in the Rain dataset

This helped me recognise that accuracy alone is misleading in imbalanced contexts. The rain project also underscored that different algorithms responded similarly to imbalance. Logistic regression underperformed, decision trees showed some improvement, and neural networks achieved the highest accuracy (~85%), but still leaned toward the majority.

Figure 10 illustrates that while neural networks performed best, the improvement over simpler models was marginal.

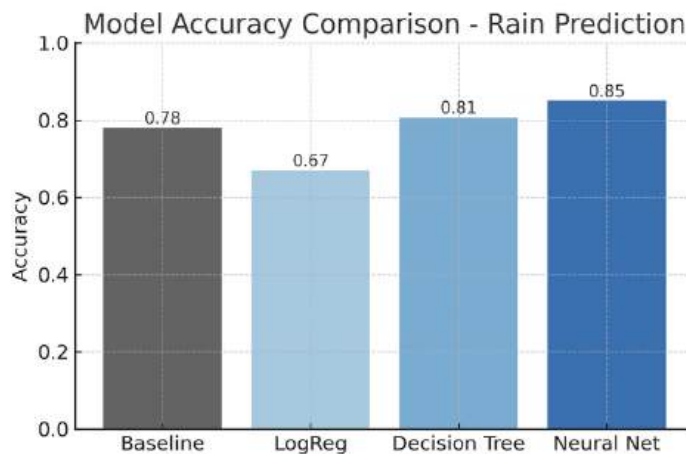


Figure 10. Accuracy comparison across models in the Rain Prediction project

This reinforced the lesson that data balance and feature engineering matter more than model complexity. Finally, data preprocessing choices, too, influence outcomes. In the rain dataset, deciding whether to impute missing features or drop them entirely significantly affected performance. This reminded me that preprocessing is just as critical as model selection.

Comparing both projects deepened the overall learning: imbalance, feature overlap, and misleading accuracy are challenges that transcend domains. By observing these issues unfold in sports and weather data, I realised that robust preprocessing, careful feature selection, and balanced data are the foundations of reliable machine learning models, far more important than the choice of algorithm alone.

8. Conclusion and Key Takeaways

This study showed that machine learning models can classify NBA players into positions with an accuracy of around 85%, but their reliability is constrained by class imbalance and overlapping feature distributions. Logistic regression and neural networks both performed well on guards and forwards but consistently struggled with centres, revealing that dataset limitations, rather than algorithm design, set the ceiling on performance.

Drawing on parallels from the rain prediction project, I learned that these challenges are not unique to sports. Across domains, majority-class dominance biases models, making overall accuracy a misleading measure of success. Evaluating models with precision, recall, and F1-scores is necessary to assess fairness across categories.

Future directions should focus on:

1. **Balancing strategies** such as oversampling, undersampling, or synthetic data generation (e.g., SMOTE).
2. **Feature enrichment**, using advanced basketball statistics (e.g., player efficiency ratings, usage rates) and refined weather features.

3. **Stronger evaluation metrics** beyond accuracy to ensure reliable performance across underrepresented groups.

In summary, this project reinforced a central lesson: data balance and feature quality matter more than algorithmic complexity. Reliable machine learning depends on preparing the data thoughtfully and evaluating it fairly — principles that hold across both sports analytics and weather forecasting.

List of Figures

Figure 1. Distribution of Height (cm) across Guards, Forwards, and Centers.

Figure 2. Distribution of Weight (kg) across Guards, Forwards, and Centers.

Figure 3. Distribution of Rebounds across Guards, Forwards, and Centers.

Figure 4. Distribution of Assists across Guards, Forwards, and Centers.

Figure 5. Distribution of Field Goal % across Guards, Forwards, and Centers.

Figure 6. Confusion Matrix for Logistic Regression baseline model.

Figure 7. Training and validation accuracy across epochs (Neural Network).

Figure 8. Training and validation loss across epochs (Neural Network).

Figure 9. Class distribution in the Rain Dataset.

Figure 10. Accuracy comparison across models in the Rain Prediction Project.

Works Cited

- “Australian Bureau of Meteorology. Rainfall and Weather Observations Dataset.” Bureau of Meteorology, Australia, n.d.
- Buda, Mateusz, et al. “A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks.” *Neural Networks*, vol. 106, 2018, pp. 249–259.
- Chawla, Nitesh V., et al. “SMOTE: Synthetic Minority Over-Sampling Technique.” *Journal of Artificial Intelligence Research*, vol. 16, 2002, pp. 321–357.
- He, Haibo, and Edward A. Garcia. “Learning from Imbalanced Data.” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, 2009, pp. 1263–1284.
- Japkowicz, Nathalie, and Shaju Stephen. “The Class Imbalance Problem: A Systematic Study.” *Intelligent Data Analysis*, vol. 6, no. 5, 2002, pp. 429–449.
- Keras API Reference. TensorFlow/Keras Developers, 2024.
- “NBA Statistics Database.” NBA.com, n.d.
- Scikit-learn: Machine Learning in Python. Scikit-learn Developers, 2024.

Early Signs of Autism in Children By Aadhya Rakesh

Abstract

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by differences in social communication, behavior, and sensory processing. Early detection is critical, as intervention programs implemented before the age of three can significantly improve developmental outcomes. This paper reviews existing literature on the early signs of autism in children, focusing on social communication differences, repetitive behaviors, and sensory sensitivities. In addition, a brief survey was conducted to examine parental perceptions of when early signs of autism first appear. The results revealed a clear gap in awareness: parents of children with ASD were more likely to recognize that signs can emerge before age two, while parents without direct experience often identified later ages or expressed uncertainty. These findings highlight a lack of widespread public understanding of early developmental indicators. Therefore, this paper underscores not only the importance of early identification but also the urgent need for increased public education and awareness to promote timely diagnosis and intervention.

Introduction

Autism Spectrum Disorder (ASD) affects approximately 1 in 36 children in the United States (Centers for Disease Control and Prevention [CDC], 2023). It is marked by challenges in social communication, repetitive behaviors, and restricted interests. While diagnosis can occur at any age, early identification of symptoms is critical for enabling timely interventions that improve language, cognitive, and social outcomes.

This paper examines early indicators of autism in children under the age of three, highlighting patterns in social, behavioral and sensory domains. The goal is to provide insight into early detection and encourage awareness among parents, educators and healthcare professionals. In addition to reviewing existing research, this paper also incorporates findings from a brief survey examining parental perceptions of when early signs of autism first appear. This combined approach provides both scientific and real-world perspectives on early identification.

Background / Literature Review

Early identification of ASD is supported by research indicating that children with autism often exhibit measurable differences before the age of three (Zwaigenbaum et al, 2015). The following categories summarize the most commonly observed early signs:

Social Communication Differences

- Limited Eye contact
- Lack of response to name
- Reduced social smiling or shared enjoyment

- Difficulty with joining attention (eg: pointing or showing objects)

Repetitive or Restricted Behaviors

- Hand flapping or rocking or other repetitive movements
- Insistence on routines or ritualized behaviors
- Intense focus on specific objects or topics

Sensory Sensitivities

- Overreacting or underreacting to sensory stimuli such as light, sound or touch
- Unusual interest in sensory aspects of objects (eg: spinning wheels or textures)

Early recognition of these patterns is essential as early intervention programs have been shown to significantly improve developmental outcomes in children with ASD (Rogers & Vimara, 2008). While these early indicators are well documented in research, less is known about how well they are understood by the general public, particularly among parents without direct experience with ASD.

Methods

This study combines a literature review with a brief survey to examine both established research and public perceptions regarding the early signs of ASD.

The literature review was conducted using peer reviewed journal articles, government reports and publications from reputable health organizations. Databases searched included Google Scholar, PubMed and Centers for Disease Control and Prevention (CDC) website. Sources were selected based on relevance, credibility and focus on early signs of ASD in children under three years of age.

In addition to the literature review, a survey was administered to gather data on parental awareness of early autism signs. This survey included 32 participants - consisting of 12 parents of children diagnosed with ASD and 20 parents with children without a diagnosis of ASD. Participants were asked one multiple choice question: “At what age do you think early signs of autism can first appear?” Response options included: before 12 months, 12 - 24 months, 2 - 3 years, after 3 years and Not sure.

Responses were collected anonymously and analyzed using descriptive statistics to compare trends between the two groups. The results were used to identify differences in awareness and perception related to early ASD signs. No personally identifiable information was collected and participation was voluntary.

Results

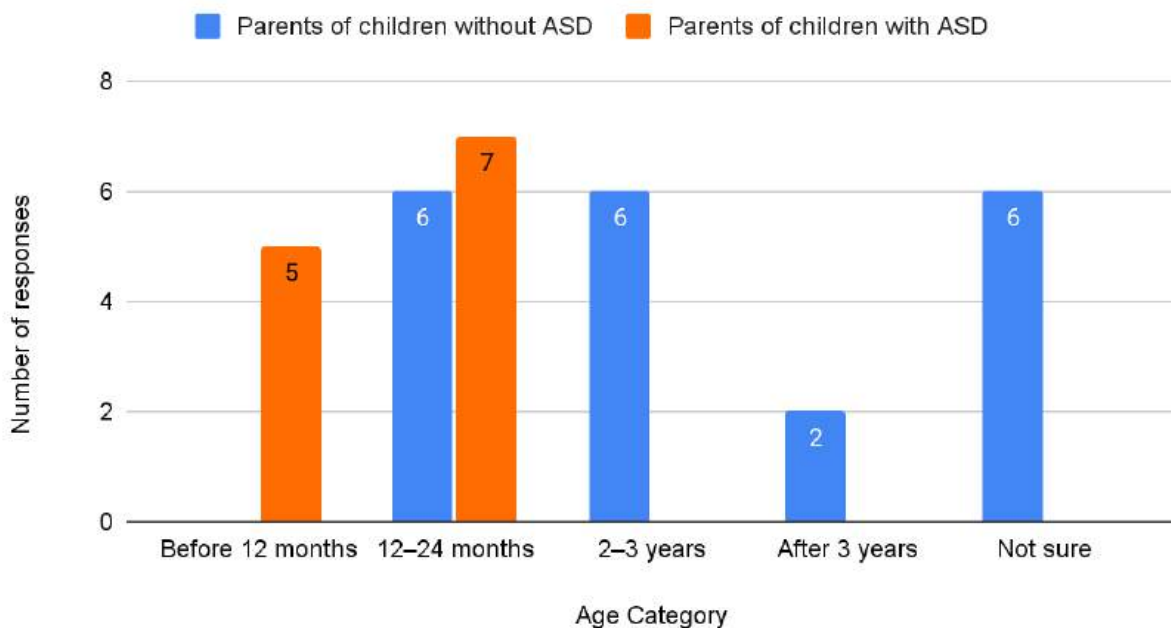
Findings from Established Research

- Social communication differences are often observable between 6-12 months, including eye contact and lack of response to name
- Repetitive behaviors and restricted interests typically emerge between 12-24 months
- Sensory sensitivities may appear in infancy but vary in intensity and presentation
- Early identification is associated with improved developmental outcomes, particularly in language, social interaction and adaptive functioning.

Findings from Public Perceptions (Survey Results)

- Among parents of children with ASD (n=12), all respondents indicated that early signs can appear before the age of two.
- Nearly half of these participants identified 'before 12 months' as the earliest point of symptom appearance.
- Among parents without children diagnosed with ASD (n=20), responses were more varied.
- A significant portion selected later age ranges, such as 2-3 years or after 3 years, while other reported uncertainty
- Overall the data indicate a clear difference in perceived timing of early autism signs between the two groups.

At what age do you think early signs of autism can first appear?



Discussion

The findings from both the literature review and the survey highlight an important gap between established research and public perception regarding the early signs of ASD. Existing

research consistently shows the early indicators of ASD can emerge within the first two years of life, with some signs observable as early as 6-12 months. These findings emphasize the importance of early detection in enabling timely intervention and improving developmental outcomes.

In contrast, the survey results reveal that awareness of these early signs is not consistent across all parents. Parents of children diagnosed with ASD demonstrated a stronger understanding of early developmental indicators, with all respondents identifying that signs can appear before the age of two and many recognizing infancy as a critical period. This alignment with established research suggests that firsthand experience with ASD may significantly increase awareness and recognition of early symptoms.

However, responses from parents without direct experience with ASD were more varied with many selecting later developmental stages or expressing uncertainty. This discrepancy indicates a potential lack of general awareness regarding the early onset of ASD symptoms. As a result, some parents may not recognize early warning signs which could contribute to delays in seeking professional evaluation and intervention.

The contrast between these two groups underscores the importance of public education and early screening efforts. Increasing awareness among parents, caregivers and educators about the early signs of ASD could lead to earlier identification and access to support services such as Applied Behavior Analysis (ABA), speech therapy etc. Tools such as early screening checklists and guidance from pediatricians can play a critical role in bridging this knowledge gap.

Overall, these findings reinforce that though scientific understanding of early ASD indicators is well established, this knowledge is not fully reflected in public perception. Addressing this gap through education and outreach is essential to ensure that more children are identified early and receive the interventions that can significantly improve their developmental outcomes.

One limitation of this study is the relatively small sample size of the survey, which may not fully represent broader public perceptions.

Conclusion

This study emphasizes the importance of early identification of Autism Spectrum Disorder (ASD) by combining established research with survey findings. Research shows that early signs can appear within the first two years of life, making early recognition critical for effective intervention and improved developmental outcomes.

The survey revealed a gap in awareness between parents with and without direct experience with ASD, with the latter group often identifying later ages or expressing uncertainty. This highlights a lack of widespread understanding of early symptoms.

Overall, these findings underscore the need for increased public education and awareness to support earlier recognition, timely intervention and better outcomes for children with ASD

Works Cited

- Centers for Disease Control and Prevention. (2023). Data & statistics on autism spectrum disorder. Data and Statistics on Autism Spectrum Disorder | Autism Spectrum Disorder (ASD) | CDC
- National Institute of Mental Health. (2022). Autism spectrum disorder: Early signs and diagnosis Autism Spectrum Disorder - National Institute of Mental Health (NIMH)
- Rogers, S.J., & Vismara, L. A. (2008). Evidence-based comprehensive treatments for early autism. *Journal of Clinical Child & Adolescent Psychology*, 37(1), 8-38. Evidence-Based Comprehensive Treatments for Early Autism
- Zwaigenbaum, L, Bauman, M. L, Choueiri, R. Kasari, C, Carter, A, Granpeesheh, D, ... Wagner, S (2015). Early identification of autism spectrum disorder. Recommendations for practice and research. *Pediatrics*, 136(Suppl 1), S10-S40 Early Identification of Autism Spectrum Disorder: Recommendations for Practice and Research - PMC

Food Quality's Effect on Overall Health in Japan By Chantal Massachi

How does food quality affect overall health in Japan? Japan's high average life expectancy points to the traditional Japanese diet and its benefits (Gabriel et al.). The Japanese diet (washoku) emphasizes eating nutrient-dense foods that contribute to better health. In addition, they implement shokuiku, a mandated food education program that promotes healthy eating habits and the importance of a balanced diet. Furthermore, kyushoku, the school meal system that provides a balanced plate, focuses on local ingredients, explores cultures, and reduces food waste. According to the FAO, a UN agency, a sustainable diet has “low environmental impacts which contribute to food and nutrition security and healthy life,” while also being accessible, economically fair, nutritionally adequate, safe, and healthy (FAO). Overall, political, social, and physical factors contribute to food-related health effects in Japan.

What political factors in Japan affect the food quality and overall health of its citizens? According to the Japanese Law Translation, run by Japan's government, it's required to promote shokuiku to educate people on gaining knowledge about "food and nutrition" and the ability to choose food for their own sake through various experiences, enabling them to adopt healthy dietary habits. Also, encouraging producers and consumers to maintain an open relationship to ensure the best-quality food (Japanese Law Translation). As mentioned, they aim to improve daily nutrition by promoting shokuiku and teaching others about the importance of quality food and its source. Furthermore, according to Hypertension Research, a peer-reviewed journal, initiatives were implemented in Japan to avoid cardiovascular diseases and reduce salt intake. The Dietary Guidelines for Japanese recommends a daily salt intake of less than 7.5g for men and 6.5g for women (Hypertension Research). Because Japan's food is high in sodium, it leads to health issues like cardiovascular disease, prompting the creation of guidelines and food education, like shokuiku, despite its traditional high-sodium foods. According to the Global Child Nutrition Foundation (GCNF), a non-profit NGO aimed at helping governments improve school meals, kyushoku in Japan began after WWII, creating the School Lunch Program Act, which requires public schools to provide safe and nutritious lunches for all students. With 99.7% of all public elementary and 98.2% of all junior high schools providing school meals, and 56% of those school meals are locally produced (GCNF). Building on shokuiku, Japan implements kyushoku in schools to improve the health and mindfulness of the Japanese. Providing high-quality food to children helps ensure a balanced diet and teaches them the importance of nourishing the body. Japan's political history has helped shape its overall health, and social factors further affect it.

How do social factors in Japan affect health from food quality? According to the OECD, a peer-reviewed international organization aiming to build better policies, with rising obesity in Japan, there also comes a rise in social pressure to be thin among young women. This leads to “nearly 30% of Japanese women between the age 15 and 19 have a BMI below 18.5” (OECD). The rise of young women becoming thinner due to societal pressures can lead them to eat less nutritious, lower-quality food, which later affects their health in the long run. Additionally, the

rise in overweight and obesity is tied to why the government is beginning to install *kyushoku* and the rise in under-eating, leading to the implementation of *shokuiku*. According to Tomoya Hanibuchi, the associate professor at Kyoto University, published in the journal “Health & Place”, community history affects social interactions; for example, traditional neighborhoods are more walkable and lead to close communities, like those found in Japan (Hanibuchi et al.). While Japan's city landscape is highly walkable, it cannot always counteract the effects of a poor diet and subpar food. However, despite rising obesity, the social ties of a neighborhood, the cultures within each community, and its walkability can lead to better health. But along with a more traditional neighborhood comes the Japanese diet with high sodium, corresponding to Hypertension Research. Furthermore, according to The Lancet, one of the most respected peer-reviewed medical journals, “disparities in socioeconomic status, material and social deprivations, poverty, and other lifestyle factors also contribute to health status variations, in Japan” (Kanchanachitra and Tangcharoensathien). With social factors ranging from marriage to economic standing, it can affect the diet of the Japanese; a poor diet and lower-quality food lead to health complications. Once again, connecting back to Japan's implementation of *shokuiku* and *kyushoku* to ensure smart, healthy eating while also making it accessible. Overall, the social environment shaped Japan's health and its physical factors because of food quality.

How do physical factors reflect the impact of food quality on health in Japan? According to the Global Nutrition Report, the world's leading independent assessment of the state of global nutrition, the physical effects of food quality in Japan has led 19% of women (15-49 years old) to be anemic, 9.5% of infants to have a low birth weight, 7.1% of children under 5 with stunted growth, 1.5% of children under 5 overweight. (Global Nutrition Report). The high anemia rate among women, showing a lack of nutrients in food, and the importance of locally sourced meals in *kyushoku*, which can help mitigate that in schools, as mentioned in GCNF. Additionally, the low birth rates and stunted growth can be caused by a lack of proper food education to pregnant women, and high-quality meals to young children and mothers. According to the OECD, the average BMI for men is 23.5 and increasing, and a steady 22.4 for women. However, being overweight is common for older people, with 25% of people over 40 having a BMI of 25 or higher (OECD). In addition to addressing obesity, Japan aims to lower the average BMI to reduce type 2 diabetes and heart disease, as seen in Hypertension Research, which points to the conclusion that part of these issues stems from high sodium intake. Specifically, because of other people having high BMI, it leads to an increased risk of getting CVD and other food-related health issues. Furthermore, according to Ana San Gabriel, the associate general manager in the Science Group's global communications department, the high sodium intake in Japanese diets leads to high blood pressure. As well as the rising obesity in men, the underweight young women, and child obesity (Gabriel et al.). This can be traced back to the loss of *washoku*, which leads to significant health problems, another reason Japan emphasizes *shokuiku*. Furthermore, the shift away from traditional Japanese food is due to the convenience of modern foods, which can lead to additional health problems, such as obesity. While *washoku* has its downsides, such as the effects of a high-sodium diet, the modern diet also has its downsides, such as cheaply

made and processed foods that lead to obesity. Overall, Japan's politics, social environment, and physical factors affect overall health.

Given the factors affecting the impact of food quality on people's health in Japan, it's essential to develop solutions to mitigate this issue. According to the EAT Lancet Commission, a global food framework as well as a peer-reviewed report from multiple perspectives, shifting agriculture to produce diverse and healthy crops to support people and the environment would improve not just people's health but also environmental stability (EAT-Lancet). By producing high-quality yields with greater diversity, it can expand people's diets and improve food quality by focusing on nutrient-rich outputs. Substituting processed, imbalanced foods with a diverse diet will help enhance nutrient intake, possibly mitigating health concerns that arise from low-quality produce. Specifically in Japan, this can help strengthen shokuiku and washoku by substituting low-quality food with high-quality, organic produce to fuel the Japanese diet. In addition to being higher-quality and more organic, food will be more flavorful, reducing the need for salt and MSG. Ultimately, leading people to live longer and healthier lives through better physical health, but it doesn't address the stigma around food, especially in teens.

Works Cited

- “Basic Act on Shokuiku (Food and Nutrition Education) - English - Japanese Law Translation.”
Www.japaneselawtranslation.go.jp, 17 June 2005,
www.japaneselawtranslation.go.jp/en/laws/view/3419/en. Accessed 9 Mar. 2026.
- FAO. “Dietary Guidelines and Sustainability.” Food and Agriculture Organization of the United Nations, 2024,
www.fao.org/nutrition/education/dietary-guidelines/background/sustainable-dietary-guidelines/en/. Accessed 15 Mar. 2026.
- Food Planet Health Healthy Diets from Sustainable Food Systems Summary Report of the EAT-Lancet Commission.
https://eatforum.org/wp-content/uploads/2025/09/EAT-Lancet_Commission_Summary_report.pdf
- Gabriel, Ana San, et al. “The Role of the Japanese Traditional Diet in Healthy and Sustainable Dietary Patterns around the World.” *Nutrients*, vol. 10, no. 2, 3 Feb. 2018, p. 173,
www.ncbi.nlm.nih.gov/pmc/articles/PMC5852749/, <https://doi.org/10.3390/nu10020173>. Accessed 9 Mar. 2026.
- Global Child Nutrition Foundation. “Shokuiku – How Japan Leverages School Meals as a “Living Textbook” for Lifelong Healthy Eating · GCNF.” GCNF, 26 June 2024,
gcnf.org/how-japan-leverages-school-meals-as-a-living-textbook-for-lifelong-healthy-eating/. Accessed 9 Mar. 2026.
- Global Nutrition Report. “Japan Nutrition Profile - Global Nutrition Report.” Global Nutrition Report, 17 Feb. 2020,
globalnutritionreport.org/resources/nutrition-profiles/asia/eastern-asia/japan/. Accessed 9 Mar. 2026.
- Hanibuchi, Tomoya, et al. “Does Walkable Mean Sociable? Neighborhood Determinants of Social Capital among Older Adults in Japan.” *Health & Place*, vol. 18, no. 2, Mar. 2012, pp. 229–239, <https://doi.org/10.1016/j.healthplace.2011.09.015>. Accessed 9 Mar. 2026.
- Ikeda, Nayu, et al. “Evaluation of Public Health and Economic Impacts of Dietary Salt Reduction Initiatives on Social Security Expenditures for Cardiovascular Disease Control in Japan.” *Hypertension Research*, vol. 48, 17 Feb. 2025,
<https://doi.org/10.1038/s41440-025-02108-z>
- Kanchanachitra, Churnrurtai, and Viroj Tangcharoensathien. “Health Inequality across Prefectures in Japan.” *The Lancet*, vol. 390, no. 10101, Sept. 2017, pp. 1471–1473,
[https://doi.org/10.1016/s0140-6736\(17\)31792-0](https://doi.org/10.1016/s0140-6736(17)31792-0).
- OECD (2019), *OECD Reviews of Public Health: Japan: A Healthier Tomorrow*, OECD Publishing, Paris. <https://doi.org/10.1787/9789264311602-en>

The Academic Outcomes of Student-Athletes: A Correlational Study of Sports Participation and Academic Performance By Rosa Jung

Introduction

The term “student-athlete” has widely been used to represent high school students who participate in sports while also maintaining typical academic responsibilities (Sedlacek & Adams-Gaston, 1992). Participation in athletics during high school often requires significant commitments of time and energy, creating challenges in balancing sports with academic expectations. Because of these demands, questions have arisen regarding the effect that sports participation may have on students’ academic performance. In order to contribute additional insight to this discussion, this study examines the relationship between sports participation and academic performance, measured through GPA, SAT, and ACT scores, among female students currently enrolled at Emory University.

Literature Review

Introduction

In order to understand why this study focuses specifically on the effects of sports participation on GPA, SAT, and ACT scores of female student-athletes, it is crucial to understand the results of past research studies on this area. According to a meta-analysis comparing college admission tests (McManus, 1991), the most common identification of academic performance used in studies is GPA, or a student’s grade point average. Another academic indicator is the SAT, a standardized college admission test that evaluates a student’s learning capacity by focusing on educational abilities away from average high school coursework. Conversely, the ACT, or the American college test, measures what students have already learned in the high school curriculum (McManus, 1991). The use of these three different grading scales is essential to the validity of this study, as past research discerns that the use of only one grading indicator, such as only using SAT scores, may be weak predictors of college success (Ting, 2009). Furthermore, some studies suggest that physical exercise can act as effective predictors of academic achievements of high school students transitioning into college (Zhang et al., 2025). Although external variables may factor in these results, an academic study utilizing data from the official College Board found that participating in sports is correlated with higher testing scores for all students, regardless of demographics, an outside variable (Everson & Millsap, 2004).

Sports as a Stress Inducer

Research consistently shows that playing a sport significantly generates stress among high school student-athletes, highlighting stress as a leading cause of lower academic performance. Stress is induced by a variety of sources. For example, a research survey finds negative feedback and discouragement by coaches during practice and matches can result in major drops in GPA (Beron & Piquero, 2016). Additionally, having to balance academics with

sports, student-athletes tend to struggle significantly with time management. These conflicts seemingly disturb sleep quality (Chen et al., 2025), heightening a series of mental health issues including psychological stress, anxiety, and depression (Yan & Wang, 2025). Studies find that struggles with time management are derived from conflicting obligations within student-athletes' schedules (Vispo & Macaliano, 2024), which contribute to stress buildup or reduced sleep. Overall, insomnia and hypersomnia are two main effects relating to sleep quality, consequently interfering with academics (Turner II et al., 2021) and increasing stress for these students.

Another deviation of stress is contest pressure. As games become more competitive, athletes elucidate how frequent, intense training is necessary for wins (Yang et al., 2023). Specifically, an anonymous online survey focusing on sleep and mental health of elite athletes finds that injuries, overtraining, various competitions, and frequent jet lag are common contributors to disturbed sleep quality and burnout, contributing to heightened overall stress levels (Montero et al., 2024).

Benefits of Physical Activity on Academic Performance

Alternatively, evidence suggests that physical activity may benefit grades and test scores in terms of academic performance. In a study conducted in a competitive international high school in Taiwan, results show significantly greater eye movements before physical activity, suggesting that attention improves after exercise (Huang et al., 2022). Similarly, the integration of sports into instructions notably increases language proficiency, in addition to an increase in attentiveness of younger students in schools (Jiang, 2024). Moreover, incorporating athletics before taking a test or a class shows stronger improvement in both attentiveness and engagement.

Physical Activity and Positive Eating Habits

The correlation between food and fitness is greatly supported by ongoing studies. Generally, unfit people tend to have higher consumption rates of sugar and whole grains, while fitter people tend to consume more fruits and vegetables. Some food groups are shown to be associated with improved performance in different school subjects, such as language or mathematics, depending on their specific intake categories (Enriquez-del Castillo et al., 2024). Previous findings also suggest that physical fitness encourages greater academic achievement, especially among pre-pubertal schoolchildren. Specifically, the correlation between fitness and overall health shows improved attention and behavior, while positive mental health and improved self-esteem reduce stress, anxiety, and depression, positively impacting students' overall academic achievement (Hermassi et al., 2021).

High Achievement Goals of Student-Athletes

Next, playing a sport is found to increase student-athletes' achievement goals. A cross-sectional study of 64 Chinese schools examines adolescents based on their athlete engagement, finding that self-esteem is raised through cooperation of team sports that enables continuous peer interaction (Huang et al., 2025). This study indicates that self-esteem levels are

closely related to achievement goals, which many students gain when participating in a team sport. Additionally, a study consisting of 235 team sports athletes illuminates that athletes with strong competitiveness in their sports tend to strive academically (Jonker et al., 2009). These goals that student-athletes typically exhibit include desires towards competence, success, and mastery towards their sports (Lagunero et al., 2024). Additionally, research suggests that emotional achievement in student-athletes is vital to the academic performance of traditional courses including math, literature, and science (Camacho-Morles et al., 2021).

Following the developmental theory of personal growth, students involved in competitive sports in high school are typically correlated with higher grades, confidence, and achievement goals in both sporting and educational contexts (Fejgin, 1994). The cause of these factors may be related to student-athletes wanting to impress close peers. For example, positive feedback from classmates, neighbors, families, and school staff on a student playing a sport may cause sports participation to be seen as a socially acceptable justification for pursuing higher achievement goals (DeMeulenaere, 2010). A questionnaire including demographic information and cumulative GPA finds performance-oriented students to be highly motivated to prove their abilities to fellow students. The researchers conclude that these students exhibit higher participation in their sports in order to demonstrate their abilities to their peers (McCreary et al., 2019). Hence, these studies indicate that while playing a sport may increase achievement goals, it is more likely that goal-oriented students themselves are inclined to participate in school sports.

Lack of Correlation

There are a few studies that reveal no direct correlation between sports and school. Examining this topic, a meta-analysis across major high school studies finds both athletes and non-athletes to have highly similar performances in education (Byrd & Ross, 1991), suggesting that sports participation may not be a significant predictor of academic performance. Furthermore, there are instances where no adverse effects are observed at all. In fact, an ongoing prospective cohort study proclaims that physical activity shows no effect on academic achievement outcomes (Papasideris et al., 2021), contradicting earlier research that shows a strong influence of physical activity on academic achievement. Another experiment, conducted at the University of South Carolina, expresses that while being a student-athlete develops skills such as enhanced organization, problem solving, and time management, these skills are not entirely attributed to, nor enhanced through, athletic participation (Pierce, 2007).

Gender Disparities

Several studies mention differences based on gender, highlighting the importance of demographics on the overall results. An analysis of six SAT forms shows that male students perform better on geometric questions, while female students perform better on algebraic questions (Harris & Carlton, 1993). On top of these testing differences, data from the Mann-Whitney U test reveal that while male students in sporting activities experience higher

academic success than average male students, no significant differences are observed among female students who take the same tests (Pjanic et al., 2021). In terms of social acceptance and physical appearance, higher levels are distinguished among females than males, suggesting that past data may only be valid for certain genders (Ryska, 2003).

Conclusion

Past research presents varying results when comparing sports participation with academic performance among student-athletes. These studies suggest increased stress, improved focus, or even no direct relationship. Despite these findings, many existing studies contain limitations. Some rely on self-reported academic data that may introduce inaccuracies in reported scores (Cole & Gonyea, 2010). Other studies fail to control for gender disparities or inconsistent measures of academic performance. Therefore, further research is necessary to better understand how sports participation relates to academic performance among female students. In order to address all these identified gaps in this study, a research question was constructed: How do the GPA, SAT, and ACT scores of randomly selected female students currently enrolled at Emory University vary based on the number of hours per week they participated in sports during their high school senior year?

Methodology

Research Design

The predictor variable was the number of hours a student participated in sports per week, during their high school years, organized into five Likert-scale categories:

- 1) 0 hours
- 2) 1–2 hours
- 3) 3–5 hours
- 4) 6–10 hours
- 5) 10+ hours

Participants choosing 0 hours were considered non-athletes. On the other hand, participants choosing one or more hours per week were defined as student-athletes.

The criterion variables were the academic performance indicators finalized during participants' high school senior years and submitted for college, identified as:

- 1) GPA (4.0 weighted scale)
- 2) SAT scores (400-1600)
- 3) ACT scores (1-36)

These methods of academic data collection have been commonly used as academic achievement indicators in prior research examining student-athletes. Additionally, this method design aligned with previously established research on student-athletes and non-athletes enrolled

at a private liberal arts college in the Midwest. The researchers investigated differences in stress, anxiety, and social support between these students through a series of surveys including multiple questionnaires (Strauser et al., 2025). Similarly, this study utilized several questionnaires to survey students at a single institute, in order to examine the correlation between student-athletes, non-athletes, and their academic behaviors.

Participants

This study's subjects were 41 female undergraduate students currently enrolled at Emory University. Eligible participants in this study had graduated from high school within the past four academic years and had access to their high school transcript and either College Board or ACT Inc., depending on the type of college admission test they took.

While this study compared sports to academics, sports participation during participants' high school years was not a requirement since the sample included both athletes and non-athletes as a comparison. Furthermore, limiting the participants to Emory students allowed for academic consistency using the high school records that were submitted for college admission. Consequently, this subject group directly aligned with the research question, which focused on female students' high school academic performance in relation to sports participation.

Instruments

Data were collected using the Female Student Athlete Academic Survey (FSAAS), a structured questionnaire designed specifically to capture all variables relevant to the study (Appendix A). This survey included questions regarding:

- 1) Demographic information for qualification determination
- 2) Academic performance scores
- 3) Athletic history during high school
- 4) Personal open-ended questions regarding correlation perception

All items, except optional qualitative prompts, were closed-ended to allow for numerical coding and statistical analysis. Because this study was correlational, third-variable questions were included to gather participants' perspectives on external factors that may have influenced their academic performance. Using standardized academic measures strengthened the validity of the findings, while demographic and perception questions allowed for considerations of potential confounding variables. Demographic information was only shared with the researcher to check eligibility. Therefore, personal details about the students were not shared publicly.

Procedure

Data collection followed a structured process, beginning with institutional approval that was requested from Emory University by contacting the appropriate university office one month beforehand.

Once on campus, data collection took place in school-approved, high-traffic areas of Emory University. These campus areas included the campus library, the dining hall, the main café, and a residential hallway. Each site included approximately 10 participants recruited through systematic random sampling, where every third female student walking by was approached in order to increase generalizability and reduce researcher bias. Additionally, one extra participant was recruited when the researcher first arrived on campus to serve as a reserve in the case of invalid data from any other participants.

To prevent time from becoming a potential external variable, the researcher had three additional volunteers in other parts of the campus in order to conduct the survey simultaneously. Data collection started at 12:00 p.m.

Only one student was surveyed at a time. When approaching each Emory student, both the researcher and volunteers adhered to the same written script by giving a brief introduction of the study. Upon gaining consent of participation from the students, they were then handed an informed consent form (Appendix B) that contained information about the guidelines of the study. Agreeing participants signed the form and began the questionnaire. Potential participants were first briefly screened for eligibility by answering several demographic questions that were used only to determine qualification and were not included in the final data analysis. If a student did not qualify for the study, the researcher approached the next student who passed by. Because participation was entirely voluntary, students were informed that they could withdraw from the study at any time without penalty.

Participants completed the FSAAS electronically via QR code link, requiring approximately five to ten minutes to complete. During the questionnaire, the researcher answered any questions the students had.

Once all data had been gathered, responses were exported into a Google Sheets spreadsheet, accessible only to the researcher. No audio or video recordings were made during the questionnaire and all files containing data from the study were deleted after data analysis. In relation to data input:

- 1) Each row represented one participant (Subject #1–#41)
- 2) Each column represented a specific variable collected from the FSAAS, including demographic information, sport participation category, years of participation, weekly practice hours, cumulative GPA, SAT score, and ACT score
- 3) Qualitative responses from open-ended questions were placed in a separate sheet labeled “Qualitative Data” to avoid interfering with numerical analysis

Data Analysis

Data were analyzed to determine whether sport participation was associated with academic performance. Descriptive statistical methods were used to summarize data collected from the survey responses by calculating the mean, median, minimum, and maximum values of GPA, SAT, and ACT scores across different sport participation categories. Participants were also

grouped according to their reported weekly sports participation hours, and descriptive statistics were calculated for each category. The results were then organized into tables and graphs in order to visualize patterns between academic performance patterns across different levels of sports participation. Additionally, percentages were calculated to compare sports participation rates and Likert-scale responses among participants.

Qualitative responses from the open-ended questions were grouped up to be compared to the actual results of the study.

Findings/Results

Participants Demographics

Out of the 41 participants, 27 of them reported participating in at least one sport during high school, making 65.85% of them classified as student-athletes, while 34.15% of participants were classified as non-athletes. In terms of the weekly practice hours for student-athletes, the distribution of participants appeared to be skewed toward 3-5 hours or 6-10 hours (Table 1.1).

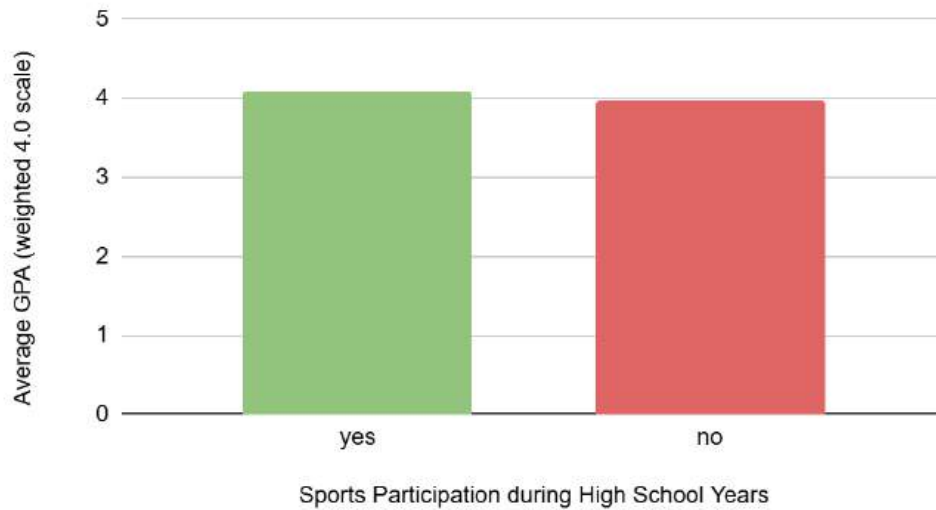
Table 1.1
Student-Athletes in Each Different Hours of Practice per Week

Practice Hours per Week	Number of Participants (N)	Percentage (of athletes)
1–2 hours	4	14.80%
3–5 hours	9	33.30%
6–10 hours	8	29.60%
10+ hours	6	22.20%
Total Athletes	27	100%

GPA

GPA responses ranged from 3.6 by a non-athlete to 5.07 by a student-athlete. The lowest GPA for student-athletes was 3.7 and the highest GPA for non-athletes was 4.5. Figure 1.1 compares the average GPAs of student-athletes and non-athletes, having 4.08 and 3.97, respectively.

Figure 1.1
Sports Participation vs. Average Weighted GPA



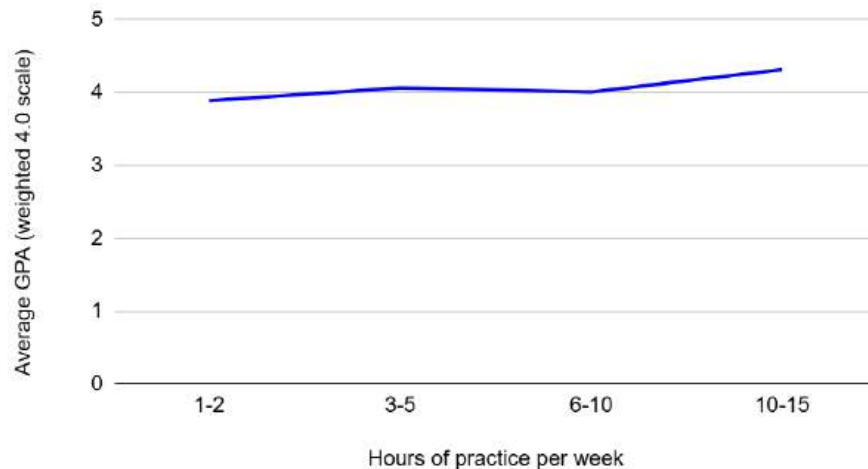
GPA responses were then compared across each weekly practice hour to determine the mean, maximum, median, and minimum values (Figure 1.2).

Figure 1.2
Relationship between Weighted GPA and Weekly Practice Hours

	Mean GPA	Maximum GPA	Median GPA	Minimum GPA
1-2 Hours of Practice	3.89	4	3.88	3.8
3-5 Hours of Practice	4.06	4.3	4	3.9
6-10 Hours of Practice	4.01	4.4	3.99	3.7
10+ Hours of Practice	4.31	5.07	4.1	4

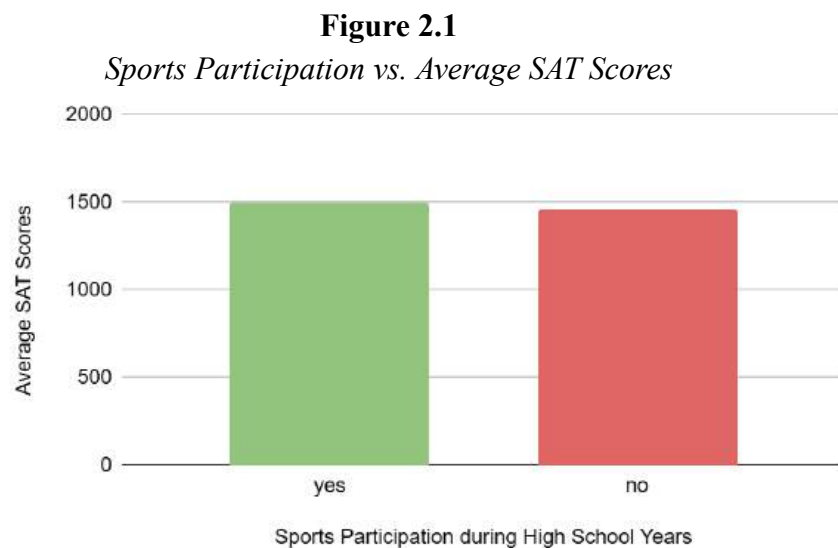
The mean GPAs of each of these weekly practice hours were graphed to show the correlation between these two variables in Figure 1.3.

Figure 1.3
Weekly Practice Hours vs. Average Weighted GPA



SAT Scores

Student-athletes had an average SAT score of approximately 1500, while non-athletes averaged approximately 1450, as shown in Figure 2.1. Although the maximum SAT score for both groups was 1550, there was a slight difference in minimum SAT scores, with non-athletes having the lowest score of 1230.



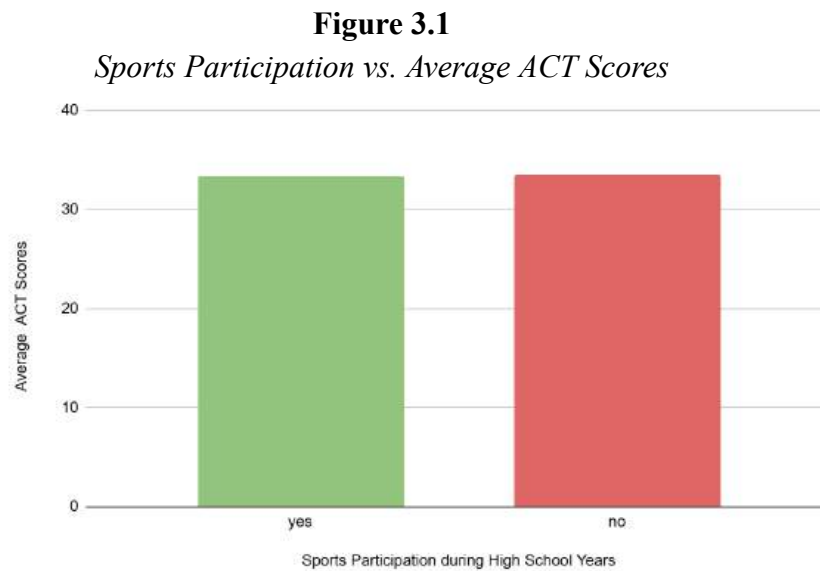
The mean, maximum, median, and minimum SAT scores of student-athletes across weekly practice hour categories were organized into Figure 2.2. Although the standard deviation of these measurements varied, the average values remained relatively constant with no significant differences.

Figure 2.2
Relationship between SAT Scores and Weekly Practice Hours

	Mean SAT Score	Maximum SAT Score	Median SAT Score	Minimum SAT Score
1-2 Hours of Practice	1510	1530	1510	1490
3-5 Hours of Practice	1492	1530	1520	1400
6-10 Hours of Practice	1505	1540	1505	1480
10+ Hours of Practice	1493.33	1520	1490	1470

ACT Scores

Student-athletes had an average ACT score of 33.38, while non-athletes had an average of 33.5, as shown in Figure 3.1. ACT scores among student-athletes ranged from 26 to 36, while scores for non-athletes had a smaller range of 33 to 34.

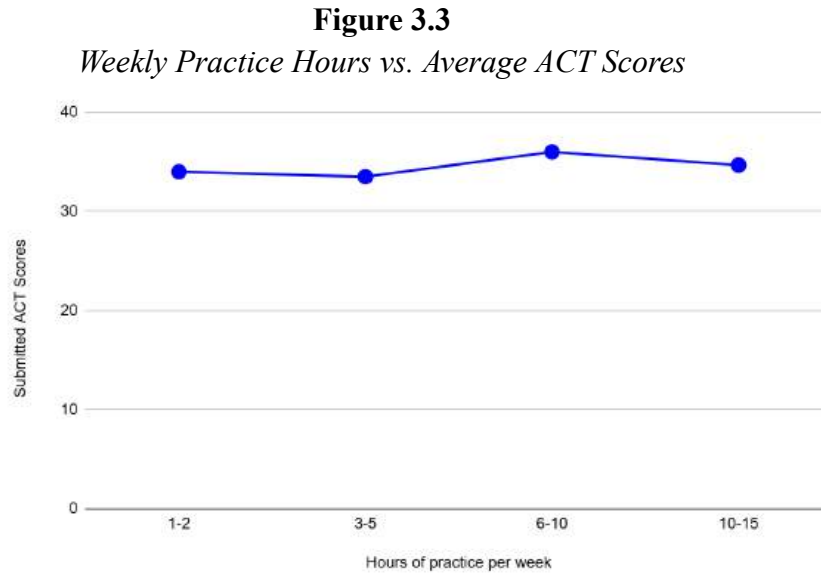


The mean, maximum, median, and minimum ACT scores of student-athletes across weekly practice hour categories were organized into Figure 3.2.

Figure 3.2
Relationship Between ACT Scores and Weekly Practice Hours

	Mean GPA	Maximum GPA	Median GPA	Minimum GPA
1-2 Hours of Practice	3.89	4	3.88	3.8
3-5 Hours of Practice	4.06	4.3	4	3.9
6-10 Hours of Practice	4.01	4.4	3.99	3.7
10+ Hours of Practice	4.31	5.07	4.1	4

Figure 3.3 compares average ACT scores and weekly practice hours for student-athletes.



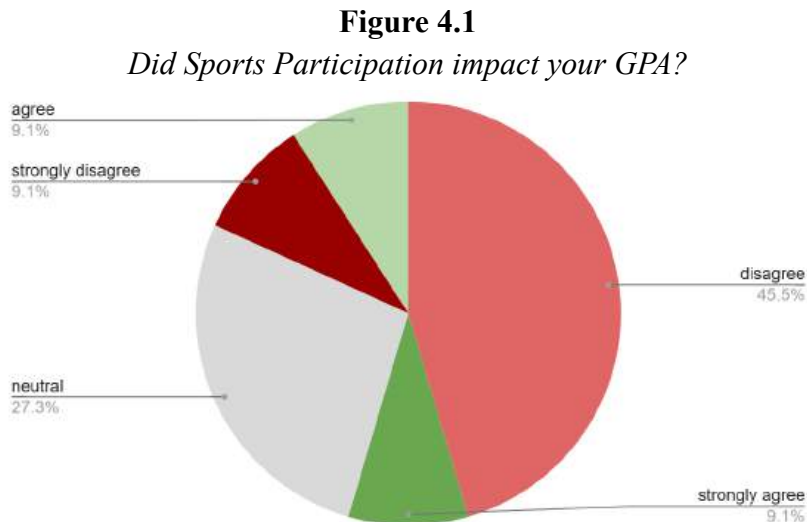
Perceived Academic Impact

Student-Athletes

Student-athletes were asked whether sports participation impacted their academic performances. Likert-scale type questions were utilized with responses ranging from strongly disagree to strongly agree.

GPA

Student-athletes were asked if sports participation had impacted their GPA (Figure 4.1). Most disagreed with this correlation, while others remained neutral in their perceptions. A smaller, equal number of the students (9.1%) either agreed, strongly agreed, or strongly disagreed.

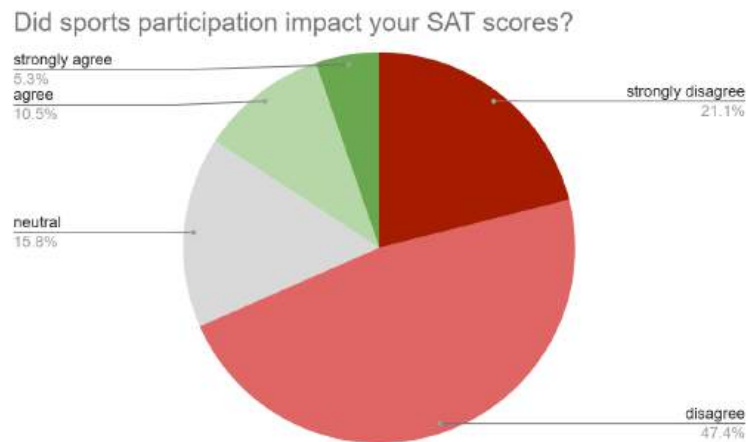


SAT Scores

When asked if sports participation impacted their SAT scores, most students disagreed, with some even strongly disagreeing. Only a small percentage of students agreed with the correlation between sports participation and SAT scores.

Figure 4.2

Did sports participation impact your SAT scores?

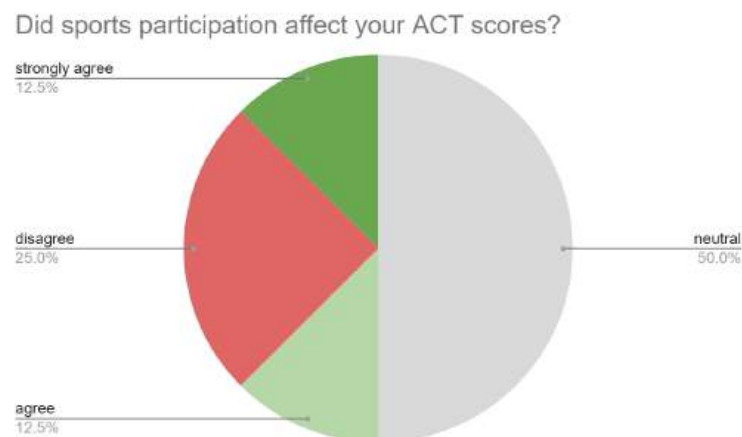


ACT Scores

Out of the 27 students that had played a sport in high school, only 8 students either took or submitted their ACT scores. These portion of students were asked if sports participation impacted their ACT scores. Figure 4.3 shows that half of the students chose neutral on the likert-scale, with an equal portion of students (12.5%) choosing agree and strongly agree. Some students disagreed with this question and there were no participants that strongly disagreed with this correlation.

Figure 4.3

Did sports participation affect your ACT scores?



Qualitative Responses

Two open-ended questions regarding student-athletes' experiences balancing sports and academics during high school were asked. For the first question, "What challenges did you face balancing academics and sports during high school," most students reported time management to be their main struggle. For example, Participant 1 stated that balancing responsibilities required "time management, completing tasks to the best of my ability." Similarly, Participants 9 and 27 both responded with "time management."

Some participants reported that sports reduced the time available for studying. Participant 8 described having "less time studying," while Participant 25 explained that "practices ran really late into the night and took away some studying time." Logistical challenges related to competitions and travel were also mentioned. Participant 14 reported difficulties due to "returning home late from matches," while participant 23 stated that "getting to the bowling alley and practicing for hours every day during the winter was hard... it took up a lot of family time and involved planning logistics to get to far away matches."

Challenges balancing sports with other responsibilities were also mentioned. For example, participant 22 reported "time struggles with balancing sport with other extracurriculars," while Participant 21 described "balancing home life and studying later at night when my parents wanted me to be in bed."

Other participants noted that sports participation did not create significant difficulties. For example, Participant 18 stated there were "none really to be honest," while Participants 10 and 40 reported "not really."

Student-athletes were also asked, "If you were not involved in sports, do you think your academic performance would have been different?" Responses varied. Some participants indicated that sports helped them focus or manage stress. For example, participant 1 stated that they were able to focus better by "sweating out study stress and playing with friends." Participant 25 similarly noted that sports "kept me disciplined... I may have done worse without them." Additionally, participant 30 stated that without sports they might be "more unfocused" because they "would have so much free time."

Others suggested their academic performance would have remained similar without sports participation. Participant 18 stated that they "don't think their performance suffered at all from playing sports."

Some participants believed their academic performance might have improved without sports due to increased available time. Participant 20 explained that without sports they would have experienced "significantly less stress, anxiety, and pressure" and would have had "more time to do work." Participant 5 similarly indicated that studying might be "easier without going to practice." Participant 22 also mentioned that sports helped them "get out stress and focus."

Discussion and Analysis

Quantitative Results

This study aimed to determine the relationship between weekly hours of high school sports participation and academic performance based on the FSAAS responses. Several trends emerged.

GPA

Figure 1.1 shows that student-athletes had a slightly higher GPA average than non-athlete students by 0.11. The small variation observed indicated that participating in athletics did not negatively affect the academic achievements of female students from Emory. Figure 1.2 also showed that the highest GPA was reported by a student-athlete, while the lowest was reported by a non-athlete, aligning with past findings that students involved in competitive sports are often correlated with higher academic outcomes (Fejgin, 1994).

When compared with different weekly practice hours, GPA remained relatively high across all hours. As exemplified by Figure 1.3, students who practiced sports for a greater number of hours per week usually maintained higher GPAs.

SAT

A similar trend was shown in SAT scores. Figure 2.1 highlights that student-athletes had a slightly higher SAT score average than non-athletes. However, when student-athletes were compared to their weekly practice hours, all values remained relatively constant.

As illustrated in Figure 2.2, SAT scores did not vary significantly between students that had different practice hours during high school, suggesting that varying levels of sports participation were not strongly associated with differences in SAT performance within this sample.

ACT

As shown in Figure 3.1, the average ACT scores between the two groups were nearly identical. Additionally, the distributions per weekly practice hours summarized in Figure 3.2 show that ACT scores remained relatively stable across different levels of sports participation. This further suggests that sports participation had little observable relationship with ACT performance among the participants in this study. However, these findings must be interpreted cautiously due to the lower number of subjects who have taken the ACT in the current study.

Self-Perception Responses

Each student-athletes' perception of the relationship between sports and academics provided valuable insights on possible external variables. Given Figures 4.1, 4.2, and 4.3, most students did not view their athletic involvement in high school as a major influence on their academic outcomes. However, the results of the study show otherwise. In fact, some student-athletes have even slightly outperformed non-athletes in their GPA and SAT scores.

The qualitative responses also added onto this insight. Most student-athletes mentioned time management as their biggest challenge in balancing sports and academics. Past studies have

also identified time constraints and conflicting responsibilities as common problems faced by student-athletes (Vispo & Macaliano, 2024).

Some student-athletes, however, indicated that athletics helped them develop beneficial skills such as discipline and focus. This response could be explained by a systematic review by the European Journal of Pediatrics that revealed physical activity interventions can improve academic performance, depending on the type of activity (Li et al., 2023).

Although sports participation may have created challenges, it seemed that most student-athletes were able to balance these challenges successfully. While most student-athletes reported challenges with balancing academics and athletics, the results actually showed that athletic participation was correlated with higher academic performance, when compared to average students.

Implications/Significance

This study's findings contribute to a broader understanding of how athletics and academics may relate to one another. Educational institutes debate on whether heavy involvement in athletics may or may not negatively affect a students' academic performance. However, this study suggests that student-athletes are generally able to maintain their academic performance at the same level as non-athletes, if not exceed them.

These findings may be especially useful for educators, parents, and guidance counselors when evaluating the role of athletics within students' education. These results indicate that sports participation does not necessarily interfere with academic achievement, and may even be correlated with strong academic performance, possibly encouraging athletic participation for students.

Limitations

While this study provides insights into the correlation between sports and academics for high school students, a few limitations must be acknowledged.

First, due to the small sample size, not all participants have taken the SATs or the ACTS. Additionally, this study only included students from a single institution, minimizing both generalizability and reliability in the findings.

Second, because the study relied on a self-reported survey, inaccurate self-reporting by the participants could be possible.

Third, the study did not control for outside variables that may have influenced academic performance. Although qualitative responses were considered, these responses only reflected participants' personal perceptions and experiences, which may not fully explain the actual causes of their academic outcomes.

Finally, sports participation in this study was measured using weekly practice hours. However, the intensity and competitiveness of different sports and schools may vary significantly. Some sports may also require more travel or time commitment than others, which could influence the student experience differently.

Future Research

Future research could expand on this study by including a larger sample of participants to allow for more diversity among the subjects studied, in order to make sure the results are generalizable to the general public. Researchers may also examine external variables, such as involvement in other extracurricular activities, in order to determine the causation of academic differences, rather than relying on a correlational analysis. Another potential direction for future research to have stronger findings may be to examine differences between levels of sports, such as team sports and individual sports, to find potential correlations between various sports. Longitudinal research following students over multiple years may also be a successful way to ethically investigate outside variables that affect student-athletes' academic performance when relating it back to their sports.

Conclusion

This study examined the correlation between different hours of weekly sports participation and academic performance, measured by GPA, SAT, and ACT scores, of the female Emory students during their high school years. Although ACT scores were not majorly impacted by weekly practice hours, both the GPA and SAT seemed to have a positive correlation with weekly practice hours, even reaching a mean that was slightly higher than those of non-athletes. The students' personal reflections about their own experience with sports revealed both problems and advantages of playing sports during high school. While there were challenges regarding poor time management skills, some reported gaining discipline and development. And although most student-athletes reported having struggled with balancing studies with practice hours, results actually showed that most participants demonstrated the ability to achieve good academic performance results, regardless of heavy involvement in sports. Hence, these findings demonstrate that dedication to sports and successful academic performance in high school can both be achieved.

Works Cited

- Beron, K. J., & Piquero, A. R. (2016). Studying the determinants of student-athlete grade point average: The roles of identity, context, and academic interests. *Social Science Quarterly* (Wiley-Blackwell), 97(2), 142-160. <https://doi.org/10.1111/ssqu.12235>
- Byrd, C. E., & Ross, S. M. (1991). The influence of participation in junior high athletics on students' attitudes and grades. *Physical Educator*, 48(4), 170-176.
- Camacho-Morles, J., Slemp, G. R., Pekrun, R., Loderer, K., Hou, H., & Oades, L. G. (2021). Activity achievement emotions and academic performance: A meta-analysis. *Educational Psychology Review*, 33(3), 1051-1095. <https://doi.org/10.1007/s10648-020-09585-3>
- Chen, F.-Y., Tsou, Y.-A., Chang, N.-J., & Chang, W.-D. (2025). Physiological parameters of sleep and the risk of obstructive sleep apnea in competitive athletes with poor sleep quality. *Life* (2075-1729), 15(4), 610. <https://doi.org/10.3390/life15040610>
- Cole, J., & Gonyea, R. (2010). Accuracy of self-reported SAT and ACT test scores: Implications for research. *Research in Higher Education*, 51(4), 305-319. <https://doi.org/10.1007/s11162-009-9160-9>
- DeMeulenaere, E. (2010). PLAYING THE game: SPORTS AS A FORCE FOR PROMOTING IMPROVED ACADEMIC PERFORMANCE FOR URBAN YOUTH. *Journal of Cultural Diversity*, 17(4), 127-135.
- Enriquez-del Castillo, L. A., Villegas-Balderrama, C. V., López-Alonzo, S. J., Flores, L. A., Martínez-Trevizo, A., & Islas-Guerra, S. A. (2024). Planetary health diet versus usual diet in adolescents. How do food and physical activity influence academic performance? *Nutrición Hospitalaria*, 41(1), 28-37. <https://doi.org/10.20960/nh.04614>
- Everson, H. T., & Millsap, R. E. (2004). Beyond individual differences: Exploring school effects on SAT scores. *Educational Psychologist*, 39(3), 157-172. https://doi.org/10.1207/s15326985ep3903_2
- Fejgin, N. (1994). Participation in high school competitive sports: A subversion of school mission or contribution to academic goals? *Sociology of Sport Journal*, 11(3), 211-230. <https://doi.org/10.1123/ssj.11.3.211>
- Harris, A. M., & Carlton, S. T. (1993). Patterns of gender differences on mathematics items on the scholastic aptitude test. *Applied Measurement in Education*, 6(2), 137. https://doi.org/10.1207/s15324818ame0602_3
- Hermassi, S., Chelly, M. S., Michalsik, L. B., Sanal, N. E. M., D. Hayes, L., & Cadenas-Sanchez, C. (2021). Relationship between fatness, physical fitness, and academic performance in normal weight and overweight schoolchild handball players in Qatar state. *PLoS ONE*, 16(2), 1-14. <https://doi.org/10.1371/journal.pone.0246476>
- Huang, H., Sha, X., Zhong, C., Ma, N., & Shui, Y. (2025). Association of football athlete engagement profiles with adolescent mental health — a latent profile analysis. *BMC Public Health*, 25(1), 1-12. <https://doi.org/10.1186/s12889-025-23239-5>

- Huang, J., Huang, H., Chang, B., & Ho, J. (2022). Evaluating the effectiveness of physical exercise in improving standardized testing performances through attention indices. *Brain & Behavior*, 12(12), 1-10. <https://doi.org/10.1002/brb3.2800>
- Jiang, N. (2024). Effects of sports games combined with English teaching on children's attention and learning performance. *Journal of Sport Psychology / Revista De Psicología Del Deporte*, 33(1), 304-311.
- Jonker, L., Elferink-Gemser, M., & Visscher, C. (2009). Talented athletes and academic achievements: A comparison over 14 years. *High Ability Studies*, 20(1), 55-64. <https://doi.org/10.1080/13598130902863691>
- Lagunero Tagare, R. J., Mancera, M., & Sumera, E. (2024). EXAMINING THE INFLUENCE OF ACHIEVEMENT GOALS IN SPORTS ON THE QUALITY OF LIFE AMONG STUDENTS IN PHILIPPINE HIGHER EDUCATION. *Revista Iberoamericana De Psicología Del Ejercicio Y El Deporte*, 19(2), 152-156.
- Li, D., Wang, D., Zou, J., Li, C., Qian, H., Yan, J., & He, Y. (2023). Effect of physical activity interventions on children's academic performance: A systematic review and meta-analysis. *European Journal of Pediatrics*, 182(8), 3587-3601. <https://doi.org/10.1007/s00431-023-05009-w>
- McCreary, C., Miller-Perrin, C., & Trammell, J. (2019). Past sports participation, self-efficacy, goal orientation, and academic achievement among college students. *Psi Chi Journal of Psychological Research*, 24(3), 194-204. <https://doi.org/10.24839/2325-7342.JN24.3.194>
- McManus, B. L. (1991). A comparison of college admission tests-the SAT and the ACT. *Clearing House*, 64(5), 339. <https://doi.org/10.1080/00098655.1991.9955886>
- Montero, A., Baranoff, J., Adams, R., & Drummond, M. (2024). Sleep and mental health of current and retired athletes across playing status and level of competition. *Sport & Exercise Psychology Review*, 19(1), 21-40. <https://doi.org/10.53841/bpssepr.2024.19.1.21>
- Papasideris, M., Leatherdale, S. T., Battista, K., & Hall, P. A. (2021). An examination of the prospective association between physical activity and academic achievement in youth at the population level. *PLoS ONE*, 16(6), 1-15. <https://doi.org/10.1371/journal.pone.0253142>
- Pierce, C. E. (2007). AN ACADEMIC SURVEY OF ENGINEERING STUDENT ATHLETES AT A DIVISION I UNIVERSITY. *College Student Journal*, 41(4), 801-812.
- Pjanić, K., Nikšić, E., & Nesimović, S. (2021). EFFECTS OF SPORTS AND MATHEMATICS EXTRACURRICULAR AND CLUB ACTIVITIES ON ACADEMIC ACHIEVEMENT OF MIDDLE SCHOOL PUPILS. *Anthropologie*, 59(1), 35-43. <https://doi.org/10.26720/anthro.20.01.29.1>
- Ryska, T. A. (2003). Sport involvement and perceived scholastic competence in student-athletes: A multivariate analysis. *International Sports Journal*, 7(1), 155.

- Sedlacek, W. E., & Adams-Gaston, J. (1992). Predicting the academic success of student-athletes using SAT and noncognitive variables. *Journal of Counseling & Development* (Wiley-Blackwell), 70(6), 724-727. <https://doi.org/10.1002/j.1556-6676.1992.tb02155.x>
- Strauser, C. M., Chavez, V., Lindsay, K. R., Figgins, M. M., & DeShaw, K. J. (2025). College student athlete versus nonathlete mental and social health factors during the covid-19 pandemic. *Journal of American College Health*, 73(1), 97-102. <https://doi.org/10.1080/07448481.2023.2201853>
- Ting, S.-M. R. (2009). Impact of noncognitive factors on first-year academic performance and persistence of NCAA division I student athletes. *Journal of Humanistic Counseling, Education & Development*, 48(2), 215-228. <https://doi.org/10.1002/j.2161-1939.2009.tb00079.x>
- Turner II, R. W., Vissa, K., Hall, C., Poling, K., Athey, A., Alfonso-Miller, P., Gehrels, J.-A., & Grandner, M. A. (2021). Sleep problems are associated with academic performance in a national sample of collegiate athletes. *Journal of American College Health*, 69(1), 74-81. <https://doi.org/10.1080/07448481.2019.1655027>
- Vispo, G. B., & Macalinao, Z. A. (2024). Extent of students involvement in extracurricular activities and academic performance among grade 8 students. *International Journal of Multidisciplinary: Applied Business & Education Research*, 5(6), 2151-2161. <https://doi.org/10.11594/ijmaber.05.06.18>
- Yan, Z., & Wang, L. (2025). The relationship between sleep disorder and mental health in athletes and its mediating role: A cross-sectional study. *PLoS ONE*, 20(3), 1-15. <https://doi.org/10.1371/journal.pone.0319813>
- Yang, M.-H., Hsueh, K.-F., Chang, C.-M., & Hsieh, H.-H. (2023). The influences of sports psychological capital to university baseball athletes' life stress and athlete burnout. *Behavioral Sciences* (2076-328X), 13(8), 617. <https://doi.org/10.3390/bs13080617>
- Zhang, B., Li, R., Liang, Y., An, Y., Zhao, C., Guo, X., & Xie, W. (2025). Physical exercise and academic achievement in chinese college students: The mediating role of self-control. *Social Behavior & Personality: An International Journal*, 53(5), 1-10. <https://doi.org/10.2224/sbp.14075>

Understanding School Bullying in China: A Multi-Factor Analysis of Risk Among Middle School Students By Yunhan Chen

Abstract

Bullying remains a serious issue among students in China, with middle school students facing especially high risks. Why does bullying occur more frequently in certain groups, and what factors increase students' vulnerability? This paper explores these questions by analyzing existing research and studies on school bullying. Drawing on published data and findings, the paper examines key factors such as gender, academic performance, family structure, boarding status, and migration background. The analysis shows that bullying is more common among boys, lower-performing students, boarding students, and those from non-traditional family backgrounds or migrant families. In addition, the research suggests that social pressure, lack of supervision, and feelings of exclusion contribute to these patterns. This paper argues that bullying is a complex issue influenced by multiple social and environmental factors, highlighting the need for greater awareness and stronger support systems in schools.

Keywords

School Bullying; China; Middle School Students; Gender Differences; Academic Performance; Family Structure; Boarding Status; Migrant Children; Hukou Status; Social Exclusion

Background

Bullying is a major problem in Chinese schools today. Over the past decade, more and more people have started to notice how often students get hurt or picked on by their peers. This usually happens in every type of school, from cities to small towns. It is not just about physical fights; it also involves emotional pain that stays with students for a long time. Even though there are now official policies from the Ministry of Education specifically to manage bullying, it is still a daily reality for many. Because of this, we need to look closely at the facts to understand why bullies keep happening.

It is important to understand that several elements of a student's living situation can contribute to prejudice. First, *hukou* (户口) is China's household registration system, which classifies people as rural or urban based on their place of registration. In some cases, local urban hukou holders have more benefits and opportunities than those with rural hukou. Second, students from non-nuclear family structures, including remarried families, foster families, or intergenerational households, face significantly higher bullying risk compared to students from traditional nuclear families. Finally, transfer students who migrate from other areas within mainland China face ostracization and isolation at their new schools.

When we look at the research data, it is clear that middle school students represent the primary demographic affected by school bullying (Li et al.). Bullying happens much more often in middle school. While some primary school studies show rates around 17.4% to 26% (Yang et

al.; Han et al.), middle school data often climbs much higher. For example, some regional studies show rates over 30% (Cheng et al.), with a sample from Xi'an reaching as high as 38.7% (Zhu and Chan, as cited in Li et al.). This is a time when kids are going through many changes and their social lives become more stressful. To understand this better, this paper will focus on five key factors: gender, boarding status, family structure, academic performance/teacher bias, and migration/hukou status.

Methodology

This study used a mixed-methods approach, combining both quantitative and qualitative analysis to better understand school bullying in China. First, a one-way ANOVA was conducted using bullying prevalence data to compare primary, middle, and high school students. The results showed statistically significant differences across school levels ($F = 4.43$, $p = 0.034$), with middle school students having a higher average rate (25.75%) than high school students ($p = 0.027$). Because middle school students show the highest bullying rates, this study focuses mainly on this group, although a small set of studies focus on a broader age range.

Second, regression analysis was used to examine how different factors, such as gender, academic performance, family structure, boarding status, and migration background, are associated with bullying involvement. Odds ratios (OR) were used to measure the level of risk for each factor.

Finally, thematic analysis was used to examine qualitative evidence, such as interviews and case studies, to help explain the patterns found in the data.

Literature review

1. Academic Performance

Academic performance is strongly correlated with involvement in school bullying. Regression analysis shows that students ranked in the bottom 30% academically have a 1.74 times higher risk of bullying involvement compared to other students (OR = 1.74; Li et al.). This suggests that lower academic achievement is a significant predictor of bullying risk.

In competitive school environments, academic ranking often determines students' social standing. Students who perform poorly in school may be more likely to be excluded, teased, or targeted by peers. As a result, lower academic performance is consistently associated with higher levels of bullying involvement.

Several social and institutional factors help explain why students with lower academic performance are more vulnerable to bullying. In many Chinese schools, academic achievement plays a major role in determining status and respect among students (Li and Hesketh). Those who struggle academically may be labeled as lazy or not smart, making them easier targets for teasing or exclusion.

Teacher attitudes can also contribute to this pattern. In highly performance-driven environments, teachers may unintentionally show more attention and support toward

high-performing students while neglecting or criticizing those who perform poorly. This halo effect toward top students can increase social pressure and reinforce negative stereotypes about low-achieving students.

As a result, students with weaker academic performance may experience lower confidence, reduced peer support, and greater exposure to bullying behavior within the school environment (Li and Hesketh).

2. Gender

Data shows a very strong link between a student's gender and their experience with bullying. In a national study of over 187,328 students in 18 Chinese cities (Qiao et al., as cited in Yang et al.), 66% of boys said they were involved in school bullying, while only 49% of girls said the same (Qiao et al., as cited in Yang et al.). This gap appears in almost every part of the country. For example, in Suzhou, about 26.4% of boys deal with bullying compared to 14.3% of girls (Wang et al.). Even in smaller cities like Jinan, the rate for boys is significantly higher at a percentage of 16.4% compared to 12.7% for girls (Yang et al.).

Girls are more likely to deal with relational bullying, such as being ignored by friends or having rumors spread about them. Boys also face more attacks online, which is called cyberbullying.

Some researchers argue that female students are more likely to be targeted through relational and psychological bullying. This specific form of abuse includes being ignored by peers, intentional social exclusion, or having malicious rumors spread about them. In some studies, girls reported slightly higher rates of social isolation (10.4%) compared to boys (8.4%) (Li et al.). However, this counterargument is outweighed by the fact that boys face a much higher total risk across almost all categories of bullying. National data involving over 187,000 students shows that 66% of boys report involvement in bullying, compared to only 49% of girls (Qiao et al.). Regression analysis further confirms that boys are 1.6 times more likely to be victims and 1.9 times more likely to be perpetrators overall (Han et al.). While girls may experience more social exclusion, the severity of physical violence is much higher among male students. For instance, boys are significantly more likely to report being "hit, kicked, or pushed" than their female classmates (Cheng et al.; Li et al.). Therefore, although relational bullying is a serious issue for girls, the frequency, physical danger, and broader prevalence of school bullying remain significantly more severe for boys.

According to regression analysis, boys are 1.6 times more likely to be victims and 1.9 times more likely to be the ones who bully others (Han et al.). This is true even for migrant children in cities, where the rate for boys is 32.4% compared to 15.5% for girls (Yang et al.). Whether it is in Beijing or Inner Mongolia 20.41% (Zhang et al.), the numbers consistently tell

us that boys face a much higher risk.

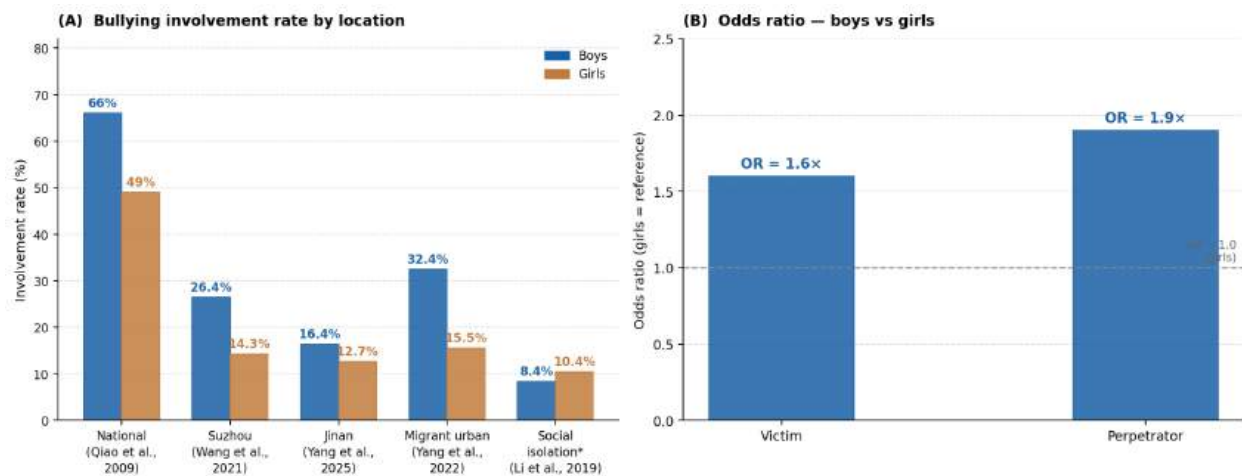


Figure 1. Gender differences in school bullying involvement among Chinese students. (A) Bullying involvement rates (%) across five samples. The national sample covers 187,328 students across 18 Chinese cities (Qiao et al., as cited in Yang et al.). Regional data include Suzhou (Wang et al.) and Jinan (Yang et al.). The migrant urban subsample is drawn from Yang et al.. *Social isolation is the only bullying subtype in which girls (10.4%) report a slightly higher rate than boys (8.4%) (Li et al.). (B) Odds ratios from regression analysis (Han et al.), with girls as the reference group (OR = 1.0). Boys are 1.6 times more likely to be victims and 1.9 times more likely to be perpetrators of bullying than girls. The dashed line indicates the reference value. *Figure generated by author.*

There are several clear reasons why boys in middle school get into fights more often. Each gender has a very different kind of bullies. Boys usually face physical bullies, like being kicked, pushed, or locked in a room. In one study analyzing school violence in a Chinese city, the detection rate for physical violence against boys was 70.8%, while girls had a much lower percentage of 48.3%. The researchers interviewed one 14-year-old girl who explained the difference in styles, noting, “Schools don’t allow fighting, so there isn’t much physical bullying. When fighting occurred, our informants said this was always among boys” (Li and Hesketh 164).

Many researchers believe this pattern exists because of gender norms in society. Often, people are more forgiving of aggressive behavior in boys because they see it as a sign of being manly (Huang et al.; Fei et al.; Yang et al.). As researchers have noted, “[D]ominant conceptions of manhood may lead to tolerance of boys acting out expressions of aggression, violence, sexual power, and homophobia” (Chen et al. 114). In middle school, boys often try to show their power or gain a high social status by acting tough or joining gangs to build “masculine capital” (Wang et al.).

Consequently, boys who show behaviors that do not match traditional masculine expectations are often targeted for bullying. Teachers and parents frequently notice that boys solve problems directly with their fists, while girls’ conflicts are more hidden and social. In

extreme cases, bullying and aggressive peer dynamics can even lead to serious violence. For example, in 2024 a widely reported case in Handan, Hebei Province involved three middle school boys who murdered their 13-year-old classmate after long-term bullying. According to reports, the victim had been repeatedly bullied by the suspects at school before the incident. The three boys later attacked the victim with a shovel and buried the body in an abandoned greenhouse. The case shocked the public and sparked national discussions about school bullying and juvenile violence in China. This example shows how persistent bullying and violent peer culture can escalate into extremely serious consequences.

3. Family Structure

Family structure is also an important factor related to bullying involvement. Research shows that students from non-nuclear families experience a 1.55 to 2.60 times higher risk level of bullying involvement (Zhang et al.). In addition, so-called “left-behind children,” whose parents migrate to large cities like Shanghai and Beijing for work, which leads to less communication with them, report higher victimization rates than students living with their parents (Zhang et al.). Studies show that 16.3% of left-behind middle school children report bullying victimization compared to 12.1% of non-left-behind children (Feng et al.). Within that situation, students that are left behind will experience isolation and separation from their parents and school, which often leads to bullying and depression.

These findings indicate that family environment and parental presence are important factors influencing students’ vulnerability to bullying.

Several social mechanisms may explain why students from non-nuclear families experience higher bullying risk. One major factor is the absence of stable parental support and supervision. When children grow up without consistent parental presence, they may lack emotional guidance and protection from difficult social situations at school.

For example, left-behind children often live with grandparents or other relatives while their parents work in distant cities. This situation may reduce daily communication and emotional support from parents. As a result, these students may have fewer resources to cope with peer conflicts or bullying.

In addition, family instability can sometimes affect a child’s confidence and social relationships. Students who experience family stress or limited supervision may become easier targets for peer abuse or social exclusion within the school environment (Feng et al.; Liu et al.).

4. Boarding Status

Data shows a strong relationship between boarding status and bullying involvement. Students who live in school dormitories face a significantly higher risk compared to those who commute. Research shows that boarding students are more likely to be victims of traditional bullying, with a risk of 1.7 times higher than non-boarding students (OR = 1.7, 95% CI: 1.4–2.1; Li et al.).

In terms of prevalence, one study of 10,528 rural primary school students found that 83% of boarding students reported being bullied at least once per month or week, compared to 72% of non-boarding students (Wang et al.). Another study shows that 67% of boarding students reported being bullied, which is higher than 61% for non-boarding students (Liu et al.). Cumulatively, the data shows that boarding students have higher risks of both victimization and perpetration.

There are several reasons why boarding students face higher bullying risks. One important factor is the closed living environment. Boarding students spend almost 24 hours a day with their peers, which increases the chances of conflict. Bullying can also extend from the classroom into the dormitory. For example, one student reported that during rest time, a roommate kept talking, and after repeated reminders, the conflict escalated into a physical fight (Wang et al.).

Another key reason is the lack of supervision. Dormitories are often seen as a “supervision vacuum,” where teachers are not always present. This makes it easier for bullying to happen without being noticed or stopped. One interviewee noted that while cameras were installed in classrooms and hallways, areas like dormitories and playgrounds had no monitoring, making them common locations for bullying (Li and Hesketh).

Boarding students also have nowhere to hide. Unlike commuting students who can return home after school, boarding students must continue living with their peers, including those who bully them. This makes bullying more continuous and harder to escape.

In addition, boarding students are often separated from their families. This weakens their support system, making it harder for them to cope with bullying. This issue is more serious in rural areas, where 77% of students live in dormitories, compared to only 13% in urban areas. In some rural schools, students share dorm rooms with more than 10 people, which increases crowding and the likelihood of conflict.

Overall, the combination of high exposure, limited supervision, and reduced support makes boarding students more vulnerable to bullying.

5. Migration And Hukou Status

Data shows a strong link between migration status, hukou type, and students’ involvement in bullying. Migrant students are more likely to be involved in bullying compared to local students, both as victims and as perpetrators.

The cumulative bullying involvement rate for migrant students is 24.4%, based on a cross-sectional study of 7,607 students from 32 junior and senior high schools in Hunan Province, China, which is higher than 18.1% for non-migrant students. This difference appears across different types of bullying. As victims, migrant students report higher rates in all categories. For example, 17.8% of migrant students experience verbal bullying compared to 13.3% of local students. Physical bullying is reported by 8.3% of migrant students versus 4.2% of local students (Yang et al.).

Migrant students are also more likely to be perpetrators. Their physical bullying rate is 4.5%, compared to 1.6% for non-migrant students, and their verbal bullying rate is 6.4%, compared to 3.9%. In addition, regression analysis shows that students with a rural hukou are 1.31 times more likely to become perpetrators (OR = 1.31). Being from a migrant family is also a significant predictor of being both a bully and a victim, with an odds ratio of 1.23 (Wang et al.).

Broadly speaking, migrant students have higher risks in both roles, but the victim rate is higher than the perpetrator rate, showing that they are more often targets of bullying.

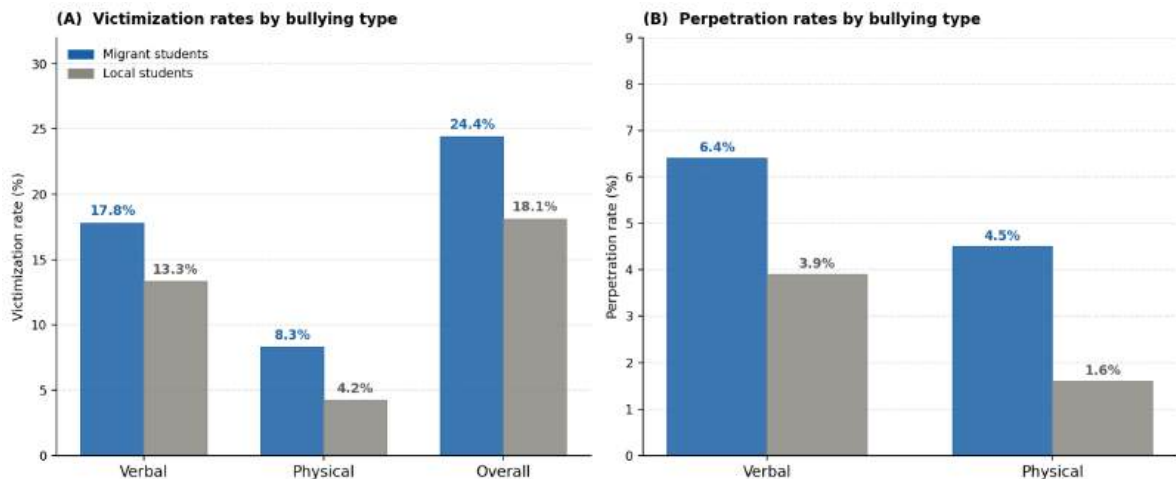


Figure 2. Bullying involvement rates by migration status among Chinese students. (A) Victimization rates (%) broken down by bullying type. Migrant students report higher rates than local students across all categories, including verbal bullying (17.8% vs. 13.3%), physical bullying (8.3% vs. 4.2%), and overall victimization (24.4% vs. 18.1%). (B) Perpetration rates (%) by bullying type. Migrant students also show higher perpetration rates for both verbal bullying (6.4% vs. 3.9%) and physical bullying (4.5% vs. 1.6%). Data are drawn from a cross-sectional study of 7,607 students across 32 junior and senior high schools in Hunan Province (Yang et al.). *Figure generated by author.*

There are several reasons why migrant students are more likely to be both victims and perpetrators. One important reason is their identity as outsiders. Migrant students are often seen as “outsiders” in urban schools and may be excluded because of differences in dialect and habits. This makes them more likely to be bullied by local students. For example, one migrant student reported being mocked and excluded by classmates after transferring to a city school, including having tadpoles placed in his water bottle and his secrets spread for ridicule. He also felt ignored by teachers, who rarely supported migrant students (Li and Hesketh).

Family background also plays a role. Migrant families usually move to cities for work, and the economic and living pressures in large cities may be passed down to children. Parents often have lower levels of education and may not fully understand their children’s needs, and they are usually busy working, which means children receive less support and protection at school.

At the same time, some migrant students may become perpetrators as a way to cope with stress. Bullying can be used as a way to release pressure or gain attention. In some cases, especially among boys, bullying is seen as a way to gain respect and be accepted by peers.

Many migrant students also fall into the category of bully-victims. They are bullied because of their background and then respond with aggression. Some have experienced physical punishment at home, which increases their risk of bullying involvement. Research shows that migrant children who have been physically punished by parents or guardians are 2.99 times more likely to be involved in bullying (OR = 2.99; Yang et al.). In addition, lack of supervision may lead them to join negative peer groups, increasing the chance of both being bullied and bullying others.

In extreme cases, this cycle can escalate into serious violence. One case involved a migrant student in Hunan Province on May 17, 2019, who was repeatedly bullied and seen as an “easy target outsider.” After earlier attempts to seek help from teachers failed, he was later surrounded and attacked by a group of classmates. In a state of fear and desperation, he fought back with a knife and seriously injured two students. The court later ruled his actions as self-defense (Shengqiao Lin), showing how the failure of school protection can push victims to become perpetrators.

Limitations

This study has several limitations. First, as a literature review, this study is based on existing research rather than original data, so it cannot independently investigate the specific nature of cause-and-effect relationships. For example, it is unclear whether poor academic performance leads to bullying or if being bullied affects students’ performance. Future research could use longitudinal studies to better understand this.

Second, much of the data comes from self-reported studies, which may be affected by bias. Students may underreport bullying due to fear or social pressure.

In addition, there are gaps in the existing research. Some studies focus on specific regions or use relatively small sample sizes, which may limit how well the findings represent all students in China. Some data is not recent (the oldest data is from 2009), so it may not fully reflect current trends.

Finally, this paper does not include all possible factors, such as personality, school climate, and online environments, which may influence bullying. Instead, it uses ANOVA to compare bullying prevalence across school levels.

Conclusion

In conclusion, school bullying in China is affected by several key factors, including gender, academic performance, family structure, boarding status, and migration background. These factors influence students’ daily lives and how they treat others, which increases the chance of being involved in bullying. This shows that bullying is not caused by just one reason but rather by many different interconnected social and environmental factors.

Among these factors, academic performance and family structure were revealed to be the most important. In many schools, grades affect how students are treated. Students with low grades are more likely to be left out or looked down on. At the same time, students without strong family support may feel more alone and have fewer people to help them when they face bullying.

Given the findings of this study, schools and families should do more to prevent bullying. Teachers should build better values in the classroom and not judge students only by their grades or encourage others to avoid low-performing students. Parents should dedicate greater attention to their children's mental health. For boarding students, it is important for a trusted adult to check their emotional condition more often. Students should also learn not to treat others differently because of stigmatized elements of their background, such as family situations or hukou.

These factors represent a serious and meaningful component of China's middle school bullying crisis, and action is needed in order to protect students.

Works Cited

- Chen, Binli, et al. "Does Friend Support Matter? The Association between Gender Role Attitudes and School Bullying among Male Adolescents in China." *Children*, vol. 9, no. 8, 29 July 2022, p. 1139, <https://doi.org/10.3390/children9081139>. Accessed 30 Aug. 2022.
- Cheng, Yulan, et al. "Being Bullied and Psychosocial Adjustment among Middle School Students in China." *Journal of School Health*, vol. 80, no. 4, Apr. 2010, pp. 193–199, <https://doi.org/10.1111/j.1746-1561.2009.00486.x>. Accessed 17 July 2019.
- Fei, Liping, et al. "The Impact of Bullying Cognition on School Bullying among Chinese Primary School Students: A Moderated Mediation Model of Resilience and Sex." *Frontiers in Public Health*, vol. 12, 16 Oct. 2024, <https://doi.org/10.3389/fpubh.2024.1470322>. Accessed 27 Oct. 2024.
- Feng, Yuan, et al. "Association between Bullying Victimization and Mental Health Problems among Chinese Left-behind Children: A Cross-Sectional Study from the Adolescence Mental Health Promotion Cohort." *Frontiers in Psychiatry*, vol. 15, 7 Nov. 2024, <https://doi.org/10.3389/fpsy.2024.1440821>.
- Han, Ziqiang, et al. "School Bullying in Urban China: Prevalence and Correlation with School Climate." *International Journal of Environmental Research and Public Health*, vol. 14, no. 10, 25 Sept. 2017, p. 1116, <https://doi.org/10.3390/ijerph14101116>. Accessed 22 June 2019.
- Li, Hongjing, et al. "Analysis of the Current Situation and Factors Influencing Bullying in Junior High Schools in Backward Areas of Western, China & a Case Study of Qingyang City in Gansu." *BMC Public Health*, vol. 24, no. 1, 13 May 2024, <https://doi.org/10.1186/s12889-024-18775-5>.
- Li, Jiameng, et al. "Incidence, Risk Factors and Psychosomatic Symptoms for Traditional Bullying and Cyberbullying in Chinese Adolescents." *Children and Youth Services Review*, vol. 107, Oct. 2019, p. 104511, <https://doi.org/10.1016/j.childyouth.2019.104511>.
- Li, Jiameng, and Therese Hesketh. "Experiences and Perspectives of Traditional Bullying and Cyberbullying among Adolescents in Mainland China-Implications for Policy." *Frontiers in Psychology*, vol. 12, 6 July 2021, <https://doi.org/10.3389/fpsyg.2021.672223>.
- Shengqiao Lin. "An Exploration of the Status, Cases, and Countermeasures of Adolescent School Bullying." *Open Journal of Legal Science*, vol. 12, no. 12, 1 Jan. 2024, pp. 6947–6953, www.hanspub.org/journal/paperinformation?paperid=102377, <https://doi.org/10.12677/ojls.2024.1212986>. Accessed 13 Mar. 2025.
- Wang, Haoran, et al. "Structural Family Factors and Bullying at School: A Large Scale Investigation Based on a Chinese Adolescent Sample." *BMC Public Health*, vol. 21, no. 1, 11 Dec. 2021, pp. 1–10, eds.s.ebscohost.com/eds/pdfviewer/pdfviewer?vid=5&sid=094b0bef-d3eb-471b-a9ba-5d

- 6666146949%40redis, <https://doi.org/10.1186/s12889-021-12367-3>. Accessed 15 Apr. 2022.
- Yang, Liu, et al. “The Rising Trend of Emotional Bullying in an Eastern Chinese City: A Five-Year Epidemiological Study among School-Aged Children (2020–2024).” *Frontiers in Public Health*, vol. 13, 6 Nov. 2025, <https://doi.org/10.3389/fpubh.2025.1715737>. Accessed 3 Dec. 2025.
- Yang, Zhengmin, et al. “School Bullying among Migrant Children in China: A Cross-Sectional Study.” *Frontiers in Psychology*, vol. 13, 14 Dec. 2022, <https://doi.org/10.3389/fpsyg.2022.1027506>. Accessed 6 May 2023.
- Zhang, Xi, et al. “Prevalence and Associated Factors of School Bullying among Adolescents in Inner Mongolia, China.” *Heliyon*, vol. 10, no. 17, Sept. 2024, p. e37201, <https://doi.org/10.1016/j.heliyon.2024.e37201>. Accessed 20 Oct. 2024.
- Zhu, Yuhong, et al. “Bullying Victimization among Chinese Middle School Students: The Role of Family Violence.” *Journal of Interpersonal Violence*, vol. 33, no. 12, 27 Dec. 2015, pp. 1958–1977, <https://doi.org/10.1177/0886260515621082>. Accessed 18 Nov. 2019.

Comparing Houseplant's Potency at Improving Indoor Air Quality Against HEPA-filters in Enclosed Spaces By Hari Chimalakonda

Abstract

Indoor air pollution is a significant public health concern, with gas stoves serving as a major source of particulate matter and volatile organic compounds. This study evaluated several mitigation strategies to determine their effectiveness in improving indoor air quality following gas stove use. One control condition and six mitigation conditions were tested, with the final condition testing the combined use of a HEPA filter with the best-performing plant. AQI (air quality index), PM_{2.5} (particulate matter 2.5), PM₁₀ (particulate matter 10), formaldehyde (HCHO), and TVOC (Total Volatile Organic Compound) levels were measured throughout the experiment. Results showed that HEPA filtration was the most effective strategy for reducing AQI and PM values, while the spider plant produced the most significant improvements in HCHO and TVOC levels. When used together, HEPA filtration and spider plants significantly improved all measured air quality parameters (AQI, PM_{2.5}, PM₁₀, Formaldehyde, and TVOC). These findings suggest that combining HEPA filtration with multiple spider plants is an effective and accessible approach to improving indoor air quality in residential environments.

Introduction

(Statement of problem) How does the use of a HEPA filter and different types of plants affect the air quality in an enclosed room after the use of a gas stove?

(Hypothesis) HEPA filters will improve the PM_{2.5} and PM₁₀ levels in the kitchen after gas stove cooking more than any plant, while I predict that the Spider plant will improve air quality the most out of any plant and will improve HCHO levels and TVOC levels better than the HEPA filter.

Background Research

Indoor air pollution/quality poses a significant threat to human health. Air quality is made up of several factors. These include Particulate Matter measuring 2.5 and 10 microns, and Volatile Organic Compounds (VOCs) like formaldehyde, which make up Total Volatile Organic Compounds (TVOC). Additionally, the Air Quality Index (AQI) measures the overall presence of Particulate Matter, Ground Level Ozone, Nitrogen Dioxide, Sulfur Dioxide, and Carbon Monoxide. Since the 1700s, when people could first monitor air quality, air pollution has worsened, peaking in the late 1800s. Afterwards, air quality improved as people became more aware of the dangers. In recent years, air quality has again become a concern in some urban locations such as Los Angeles, USA, and Delhi, India (5). About 99% of people live in areas where air quality does not meet World Health Organization (WHO) standards. This leads to the deaths of around 6.7 million people every year, including from both ambient and indoor air quality (6). Outdoor or ambient air quality is considerably different from indoor air quality. Ambient air quality is influenced by industrial emissions, vehicle exhaust, natural processes, and

atmospheric conditions. Indoor air quality, in contrast, is influenced by indoor heating and cooling systems, products like paint, and appliances such as gas stoves. The World Health Organization (WHO) estimates that about 2.1 billion people worldwide have gas stoves (1). People in the United States spend about 20 hours per day indoors. It is vital to address indoor air quality.

Furthermore, over the past 40 years, clean air policies have markedly improved outdoor air quality throughout the United States. While for indoor air quality, ever since the 1970s energy crisis, more awareness has been shed on indoor air quality. Since the general public has gained increased awareness of indoor air quality, the development of new systems like natural ventilation and artificial ventilation has arisen. A significant amount of indoor air pollution comes from Particulate Matter (PM). Particulate Matter is made up of small solid particles or liquid droplets that are suspended in the air. These particles mainly consist of particles in the sizes of 2.5 or 10 micrometers and are referred to as PM_{2.5} and PM₁₀, respectively (2). VOCs are a large group of carbon-based chemicals like HCHO, also called formaldehyde, and exist as a colorless gas with a strong odor. TVOC is made up of all of the different VOC's. VOCs cause irritation in the eyes, nose, and throat, and can negatively affect some organs like the lungs. Some VOCs, like formaldehyde, are carcinogens(3). The USEPA states that an air quality level from $0 \leq 50$ is considered safe, with $50 \leq 100$ being considered unsafe for individuals with asthma or other respiratory issues. Anything above a value of 100 is considered unsafe for all individuals(4).

Some of the most well-known solutions for air quality are Electric stoves, Mechanical ventilation, and Humidifiers(3). Unfortunately, many of these solutions are not available to low-income communities. As a result, these areas have disproportionately higher levels of indoor air pollution, stemming from the use of traditional cooking methods and the lack of anything to improve the indoor air pollution.

Given these disparities, it is important to explore alternative solutions and natural approaches to improve indoor air quality. The best-known solutions are ventilation, electric stoves, HEPA filters, and plants. HEPA filters and plants are most cost-effective. Ventilation, HEPA filters, and plants are most effective (5). Most plants offer limited benefit, but spider plants provide excellent air filtration. To spread clean indoor air worldwide, the best and most cost-effective solutions are HEPA filters, ventilation, and native plants (like the spider plant from Southern Africa).

This research targets the potential of using plants to mitigate indoor air pollution, specifically in high-risk areas such as kitchens. Most times, kitchen settings are the most polluted areas of the home as the use of gas stove cooking drives the release of particulate matter and VOC's such as Formaldehyde. Many plant studies have found the general air-filtration abilities of plants, while few have tested the effectiveness of plants against indoor air pollution. This study aims to compare the effectiveness of selected houseplants and HEPA filtration in reducing indoor air pollutants following gas stove use in a closed kitchen environment. Considering the

fact that most individuals spend about 21 hours a day inside their homes, it's vital to think of indoor air quality.

To find a potential solution to this challenge, I conducted an experiment that focuses on using different air quality-improving plants, such as the spider or snake plant, to determine whether the use of plants or HEPA filters is better for indoor air quality after gas stove cooking in a closed kitchen environment. The plants being tested are the Spider plant (*Chlorophytum comosum*), Snake plant (*Dracaena trifasciata*), Boston Fern (*Nephrolepis exaltata*), Chinese Evergreen (*Aglaonema commutatum*), and the Peace Lily plant (*Spathiphyllum wallisii*). The compounds this experiment will account for are PM2.5, PM10, Formaldehyde (HCHO), and TVOC. This study will also account for the general AQI (Air Quality Index). To account for all these aspects, the experiment will test this hypothesis: HEPA filters will improve the PM2.5 and PM10 levels in the kitchen after gas stove cooking more than any plant, while I predict that the Spider plant will improve air quality the most out of any plant and will improve HCHO levels and TVOC levels better than the HEPA filter. This prediction is based on the large leaf size and correspondingly large pore size of the spider plant.

Methods and materials

- 1x gas stove
- 1x Air Quality Monitor
- 1x HEPA filter
- 3x spider plant
- 3x snake plant
- 3x Boston Fern plant
- 3x Chinese evergreen plant
- 3x peace lily plant
- 1x pen
- 1x notebook
- 1x timer

Setup and design

The entirety of this experiment was conducted within a 30 ft (Length) x 12 ft (Width) x 9.5 ft (Height) (3420 ft / 96.84362 m³) closed kitchen with a gas stove, with two doors and one window. The air quality monitor, when recording data, was placed directly next to the gas stove; this is also where the various plants and the HEPA filter were placed in trials. Baseline calibration data were recorded 10 hours after all activity had ceased in the kitchen. Every trial was conducted on different days, with no external ventilation present in the kitchen during the experiments.

Data collection

This experiment collected data on AQI (Air Quality Index), PM2.5 (Particulate Matter 2.5), PM10 (Particulate Matter 10), HCHO (Formaldehyde), and TVOC (Total Volatile Organic Compounds). The Air Quality Monitor device used to measure this data was the IGRESS formaldehyde detector/pollution tester. The gas stove was kept on the same setting for every trial, using a high flame level for approximately 20 minutes each time. All plants used in the experiment were mature, but their heights and leaf spreads varied between species. To reduce this effect, every plant group, HEPA filter, and air quality monitor was placed exactly 12 inches away from the gas stove.

Procedure(s)

NOTE: For each experimental condition (Table 1), I followed the seven shared procedural steps used for the control condition:

Steps to test the control condition:

STEP 1: Use the Air Quality Monitor and click C at the top to find calibration data using outside air and comparing it to the baseline value indoors

STEP 2: Use the gas stove for 18-20 minutes and then shut it off.

STEP 3: 1 hour after use of the gas stove, use an air quality monitor to record values for AQI, PM2.5, PM10, HCHO, and TVOC

STEP 4: 2 hours after use of the gas stove, use an air quality monitor to record values for AQI, PM2.5, PM10, HCHO, and TVOC

STEP 5: 3 hours after use of the gas stove, use an air quality monitor to record values for AQI, PM2.5, PM10, HCHO, and TVOC

STEP 6: Turn off the air quality monitor as well as the HEPA filter.

STEP 7: Repeat steps 1-6 three more times, to make a total of 4 trials, each trial being after the values from the previous trial return to baseline values

Steps to test experimental conditions

STEP 1: Use the Air Quality Monitor and click C at the top to find calibration data using outside air and comparing it to the baseline value indoors

STEP 2: Use the gas stove for 18-20 minutes and then shut it off.

STEP 3: Immediately after use of the gas stove, turn on the Levoit HEPA filter and set the time to three hours on mode 3(**Only for conditions 2 and 8**)

STEP 3: Immediately after use of the gas stove, place all 3 of the same condition-specific plants next to the gas stove(condition 3 uses snake plant, 4 uses spider plant, 5 uses Boston fern, 6 uses Chinese evergreen, and 7 uses peace lily plant). (**Only for conditions 3-8**)

STEP 4: After 1 hour of use of the gas stove, use an air quality monitor to record values for AQI, PM2.5, PM10, HCHO, and TVOC

STEP 5: After 2 hours of use of the gas stove, use an air quality monitor to record values for AQI, PM2.5, PM10, HCHO, and TVOC

STEP 6: After 3 hours of use of the gas stove, use an air quality monitor to record values for AQI, PM2.5, PM10, HCHO, and TVOC

STEP 7: Turn off the air quality monitor.

STEP 8: Repeat steps 1-, to make a total of 4 trials for the condition, each trial being after all the values from previous trials return to baseline values

For conditions using the plant-based and HEPA filtration methods, there was an additional calibration step required to set the filter and / or to place the plants.

Condition Number	Description / Filter method
1	Control
2	HEPA filter
3	Snake plant
4	Spider plant
5	Boston fern
6	Chinese evergreen
7	Peace lily
8	HEPA filter + Spider plant

Table 1 - Overview of experimental conditions

Analysis performed

Results depicted in graphs are the mean values of the collected data. Graphs were made using Google Sheets, with results in graphs including standard deviation. Results are displayed through a column chart, the results in the graph are separated based on the condition under which the data of each was collected, and the hour after the use of the gas stove, the results were found. A two-tailed t-test was used to compute P values.

Challenges and Technical Issues

Throughout this experiment, several challenges affected the consistency and accuracy of the data collected. The primary difficulty was sustaining stable environmental conditions, such as humidity and outdoor temperatures. To mitigate this, data was only recorded when humidity and air quality baseline levels were consistent with calibration values. Outside temperature variations were addressed by keeping windows closed for 24 hours prior to, and throughout each trial.

Finally, to normalize gas stove usage, all cooking sessions were conducted at a high setting for a duration of 20 minutes to minimize variability in flame intensity and pollutant output.

ANOVA Tables

AQI(**figure 1**)

Effect	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Condition	94.32	13.47	7	39.791	16.1477	p <0.001
Hour	1305.02	1305.02	1	1563.9242	1563.9242	p <0.001
Condition:Hour	33.92	4.85	7	56.0	5.8074	p <0.001

PM2.5(**figure 2**)

Effect	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Condition	7.083	1.18	7	71.421	14.7681	p <0.001
Hour	81.563	81.563	1	2115.731	2115.731	p <0.001
Condition:Hour	0.319	0.046	7	56.0	1.1834	p =0.3272

PM10(**figure 3**)

Effect	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Condition	8.259	1.18	7	71.421	14.7681	p <0.001
Hour	82.129	82.129	1	1027.9977	1027.9977	p <0.001
Condition:Hour	1.262	0.18	7	56.0	2.2561	p = 0.04278

HCHO(**figure 4**)

Effect	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Condition	5.6414e-05	8.059e-06	7	72.868	16.3555	p <0.001
Hour	0.00023066	0.00023066	1	468.1121	468.1121	p <0.001
Condition:Hour	1.2246e-05	1.749e-06	7	56.0	3.5504	p =0.003

TVOC(**figure 5**)

Effect	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Condition	0.01117	0.0015963	7	74.769	3.4821	p =0.00277
Hour	0.000841	0.000841	1	1.8307	1.8307	p =0.18148
Condition:Hour	0.005203	0.00074329	7	56.0	1.618	p =0.14932

Condition 1-8 mean AQI values with Standard Deviation

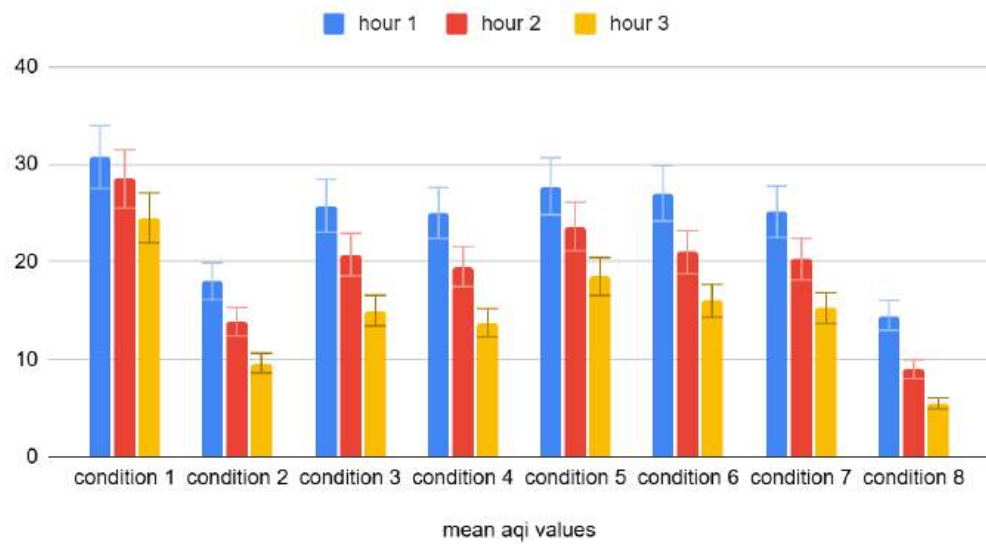


Figure 6

Condition 1-8 mean PM2.5 values with Standard Deviation

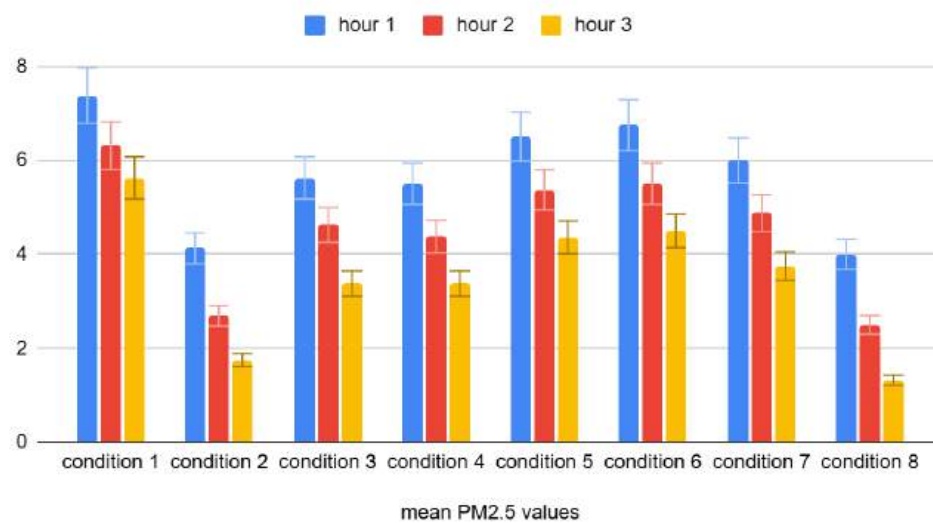


Figure 7

Condition 1-8 mean PM10 values with Standard Deviation

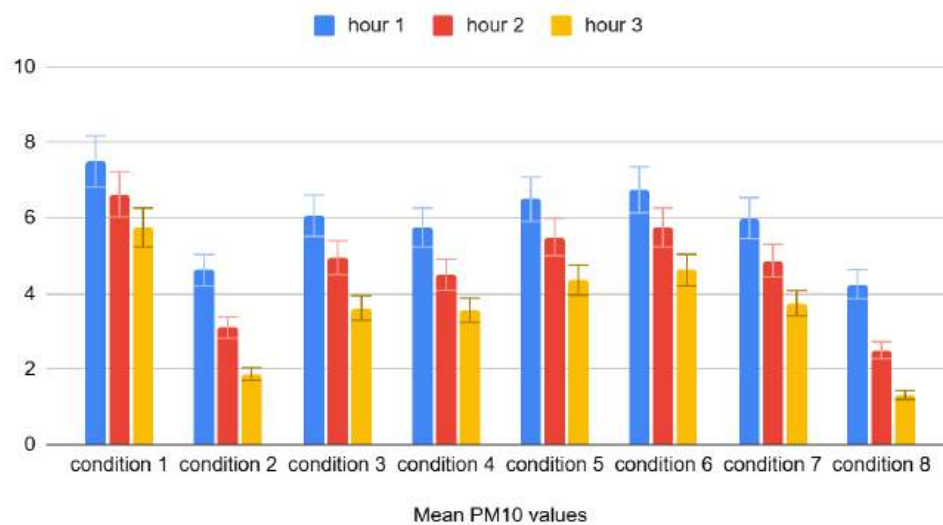


Figure 8

Condition 1-8 mean HCHO values with Standard Deviation

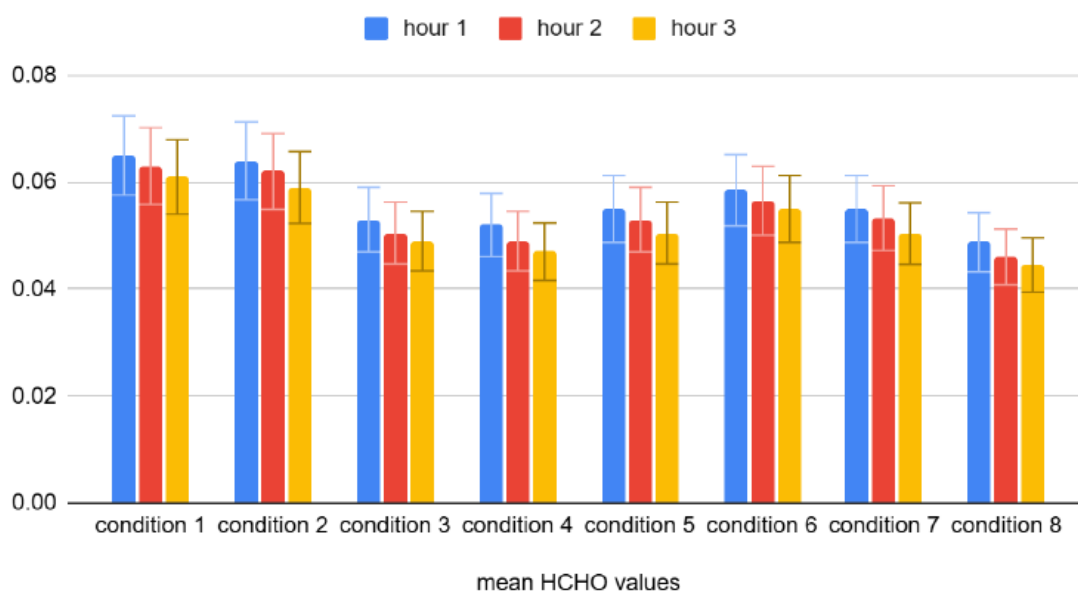


Figure 9

Condition 1-8 mean TVOC values with standard deviation

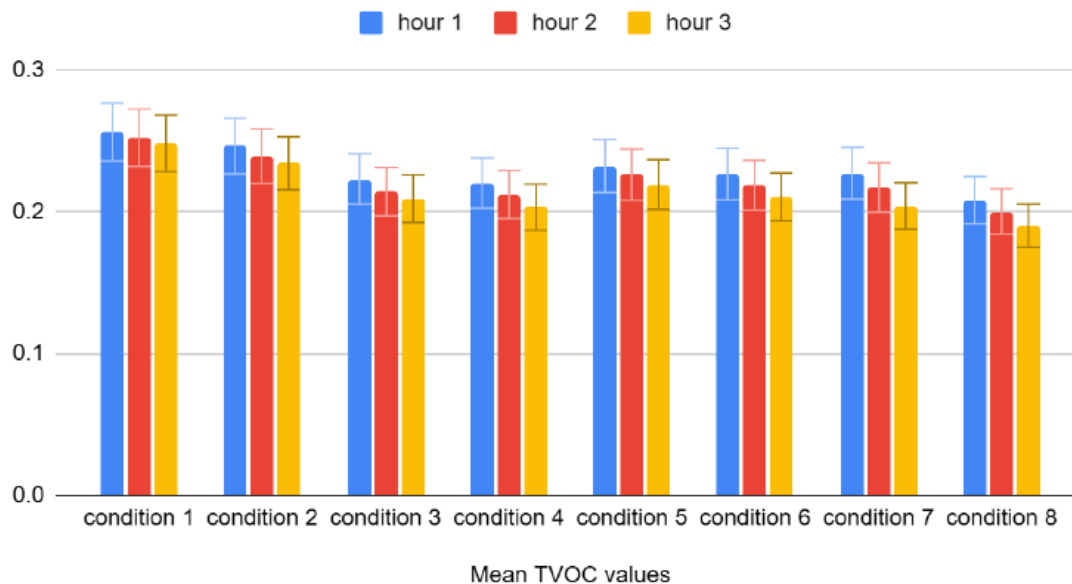


Figure 10

(Results) In this experiment, condition one served as the baseline, while the other conditions tested various solutions. The results showed that all types of air quality dropped when a gas stove was in use. Figure 6 illustrates that the HEPA filter provided the best improvement in AQI values, with the spider plant being the most effective in this area. Figures 7 and 8 indicate that HEPA filtration led to the largest reductions in PM_{2.5} and PM₁₀ levels, with the Spider Plant having the best impact among the plant options. However, in Figure 9, HEPA filters had little effect on reducing formaldehyde (HCHO) levels. In contrast, all plants significantly improved formaldehyde levels, with the Spider Plant being the most effective. Figure 10 shows that the Spider Plant had the biggest effect on TVOC values, while the HEPA filter had a slight impact on TVOC levels.

Interpretation of results and conclusion

Indoor air pollution is a major global issue that harms millions every year. Gas stoves are the top contributors to indoor air pollution, releasing harmful pollutants each time they are used. Although previous studies have looked into indoor air quality, many were limited in scope and did not provide ways to reduce pollutants. This experiment aimed to address this gap by testing specific remedies, including HEPA filters and plants like the Spider Plant, to see if they could improve indoor air quality.

The baseline calibration values recorded in this experiment were as follows: AQI=4, PM_{2.5}=1, PM₁₀=1, HCHO=0.008, and TVOC=0.033. When a gas stove was used (Condition 1), all the data values increased significantly, with a measured difference of $p = 8.0 \times 10^{-6}$.

In condition 2, placing a HEPA filter in the kitchen after using the gas stove greatly improved the AQI, PM_{2.5}, and PM₁₀ values, performing better than any plant. This supported the first part of my hypothesis. However, the HEPA filter had little effect on formaldehyde and TVOC levels in the kitchen. This aligns with how HEPA filters are designed; they trap airborne particles but are not built to handle gases like formaldehyde. Among the plants tested, the spider plant showed consistent improvement to indoor air quality, followed by the snake plant, peace lily, Boston fern, and finally, the Chinese evergreen. Interestingly, most plants did not effectively lower particulate matter levels, likely due to phytoremediation, where plants absorb gases like formaldehyde and convert them into tissue and amino acids. This process does not work as well for particulate matter. Lastly, in Condition 8, combining a HEPA filter with a spider plant resulted in better improvements across all pollutant levels than any single solution. While studies related to NASA suggested that indoor plants have limited air purification capacity in real-world settings, this research offers initial evidence that a combined approach using HEPA filtration and plants may effectively enhance air quality in a highly polluted kitchen environment.

ANOVA results showed a strong impact of Condition on AQI, along with significant effects on Hour and the Condition x Hour interaction. This indicates that different interventions not only changed overall AQI values but also affected the rate of AQI improvement over time. Similarly, Condition and Hour had a notable impact on PM_{2.5} and PM₁₀ levels. However, the Condition x Hour interaction was not significant for PM_{2.5} but was significant for PM₁₀, suggesting that while different interventions did not clearly alter the rate of PM_{2.5} improvement over time, they did alter the timeline for PM₁₀ levels. Condition significantly affected HCHO values, with notable influences from Hour and the Condition x Hour interaction. This means specific interventions substantially changed HCHO values and also significantly impacted the rate of HCHO improvement over time. For TVOC, different conditions influenced TVOC values strongly, while Hour and the Condition x Hour interaction showed no clear significance in the data. This indicates that various remedies were tied to changes in TVOC values but did not clearly affect the rate of change over time. Overall, the results suggest that using a gas stove worsens air quality.

Without any air purification device, gas stoves may pose health risks. HEPA filtration effectively addressed AQI, PM_{2.5}, and PM₁₀ levels in the kitchen but had minimal effects on HCHO and TVOC. HEPA filters can significantly improve many harmful aspects of indoor air quality. All plant types in this experiment helped enhance indoor air quality, especially regarding HCHO and TVOC levels. The Spider Plant showed the greatest overall improvement, followed by the snake plant, peace lily, Boston fern, and the Chinese evergreen. Plants serve as a versatile means to enhance air quality. In the final condition, pairing a HEPA filter with the Spider Plant led to significant improvements across all measured aspects of air quality.

Looking ahead, future experiments in indoor air quality should consider using climate controls, eliminating drafts, and regulating gas stove settings. Additionally, variables related to gas stove cooking, temperature, and humidity should be controlled. A limitation in this study was ensuring consistent gas stove usage across all trials, as variations in flame intensity or cooking

time could affect air quality. Before conducting further experiments, researchers should identify other factors influencing air quality and evaluate the effectiveness of different solutions. Future studies could explore how electric stoves affect indoor air quality compared to gas stoves. Another possible area of research could assess the impacts of genetically modified plants on indoor air quality versus non-modified plants. If an air purifying device is used in future studies, following proper cleaning instructions would be essential; similarly, watering plants according to their needs would enhance the reliability of experimental outcomes.

In conclusion, this study offers initial support for the hypothesis that both HEPA filtration and indoor plants can improve air quality in this experimental setup, though they target different pollutants. The findings highlight that using gas stoves negatively affects a home's air quality. This study concludes that the best way to improve indoor air quality is through a combination of artificial solutions, like a HEPA filter, along with natural solutions, such as the Spider Plant (*Chlorophytum comosum*).

Limitations

The experiment relied on a single air quality monitor (IGRESS formaldehyde detector), which had a detection limit of 0.005 PPM (Parts Per Million) and was calibrated once before each trial. However, like all consumer monitors, readings may come with measurement errors and may not accurately reflect changes in pollutant levels. Since each condition was only tested four times, the sample size was small ($n=4$), which limits the confidence in generalizing results beyond a kitchen setting. Despite efforts to control outside factors like plant size, humidity, and airflow, environmental changes, such as outdoor temperature, should be more strictly managed in future studies.

Acknowledgements

Special thanks to my family and mentor, who continuously offered guidance and support for my project.

Works Cited

- World Health Organization. "What Are the WHO Air Quality Guidelines?" World Health Organization, 22 Sept. 2021, www.who.int/news-room/feature-stories/detail/what-are-the-who-air-quality-guidelines.
- Zhang, Ling, et al. "Indoor Particulate Matter in Urban Households: Sources, Pathways, Characteristics, Health Effects, and Exposure Mitigation." *International Journal of Environmental Research and Public Health*, vol. 18, no. 21, 21 Oct. 2021, p. 11055, <https://doi.org/10.3390/ijerph182111055>.
- American Lung Association. "Volatile Organic Compounds." [Www.lung.org](http://www.lung.org), 2 Nov. 2023, www.lung.org/clean-air/indoor-air/indoor-air-pollutants/volatile-organic-compounds.
- AirNow. "Air Quality Index (AQI) Basics." AirNow, 2025, www.airnow.gov/aqi/aqi-basics/.
- US EPA. "Improving Indoor Air Quality | US EPA." US EPA, 25 Oct. 2018, www.epa.gov/indoor-air-quality-iaq/improving-indoor-air-quality.
- "Air Pollution, London vs. Delhi." Our World in Data, ourworldindata.org/grapher/air-pollution-london-vs-delhi.
- World Health Organization. "Air Pollution Data Portal." [Www.who.int](http://www.who.int), World Health Organization, 2024, www.who.int/data/gho/data/themes/air-pollution.
- Lowther, Scott D., et al. "How Efficiently Can HEPA Purifiers Remove Priority Fine and Ultrafine Particles from Indoor Air?" *Environment International*, vol. 144, 1 Nov. 2020, p. 106001, pubmed.ncbi.nlm.nih.gov/32739515/, <https://doi.org/10.1016/j.envint.2020.106001>.