IJHSR

International Journal of High School Research

February 2025 | Volume 7 | Issue 10
ijhsr.terrajournals.org
ISSN (Print) 2642-1046
ISSN (Online) 2642-1054

Marine Biology Research at Bahamas

Unique and exclusive partnership with the Gerace Research Center (GRC) in San Salvador, Bahamas to offer marine biology research opportunities for high school teachers and students.

- Terra has exclusive rights to offer the program to high school teachers and students around world.
- All trips entail extensive snorkeling in Bahamian reefs as well as other scientific and cultural activities.
- Terra will schedule the program with GRC and book the flights from US to the GRC site.
- Fees include travel within the US to Island, lodging, meals, and hotels for transfers, and courses.
- For more information, please visit terraed.org/bahamas.html

Terra is a N.Y. based 501.c.3 non-profit organization
dedicated for improving K-16 education

Table of Contents

October 2025| Volume 7| Issue 10

Editorial Board

International Journal of High School Research

IJHSR

ijhighschoolresearch.org

RESEARCH ARTICLE

CAR-T Cells: Novel Designs and Emerging Therapeutics for B-Cell Malignancies

Alex Zhang

Fraser Heights Secondary School, 16060 108 Ave, V4N 1M1, Canada; zhangj.alex@gmail.com

ABSTRACT: Chimeric antigen receptor (CAR)-T cell therapy is a novel approach to treating hematologic cancers such as leukemia. This review paper focuses on optimization strategies for CAR-T cell engineering to improve therapeutic durability and reduce relapse rates. Key methods discussed include advances in dual antigen targeting to mitigate antigen escape alongside signal enhancements in the costimulatory domain to enhance T cell persistence. Next-generation CAR-T cells will be discussed and further address challenges like tumour microenvironments. By incorporating these advancements in dual antigen targeting, combined with innovations in CAR cell body design and development processes, CAR-T cell therapy has the potential to significantly improve therapeutic efficacy, reduce complications, and increase accessibility for leukemia patients.

KEYWORDS: Cellular and Molecular Biology, Genetics, Chimeric Antigen Receptor (CAR), CAR-T Cell Therapy, B-Cell Malignancies, Leukemia, CD19, CD22, CD37.

Introduction

Cancer is a life-changing topic and an obstacle to many families worldwide. Cancer, with its complex nature in biology making it challenging to treat effectively, is one of the leading causes of mortality and morbidity globally: in 112 out of 183 countries, estimates from the WHO in 2019 rank cancer as the first or second leading cause of death.[1] Cancer cases are expected to be 28.4 million in 2040, a 47% rise from 2020, which has been a burden on developing countries due to demographic changes.[1] A seemingly death-bound disease has caused concern over an improved treatment or cure.[2] Modern cancer treatment approaches have had varying degrees of success but are frequently associated with a variety of side effects, like targeting other cells that are not cancer cells, harming the patient. [3] Radiotherapy, one of the most common cancer treatments, could result in radio-resistance;[4] hadron therapy, a new alternative involving protons and light nuclei, has no insight to its long term impacts;[5] chemotherapy, based on cytotoxic drugs, often results in a impaired quality of life and severe side effects;4 surgery, effective but invasive; and other treatments all have led to certain negative downsides.[6] Individualized therapy offers a promising solution by leveraging a patient's immune system to eliminate cancer cells with high specificity to reduce off-target effects and to provide more personalized and potentially less harmful treatment options.[7] Despite all this, recently, a viable and new approach to cancer has been revealed, chimeric antigen receptor (CAR)-T cell therapy.

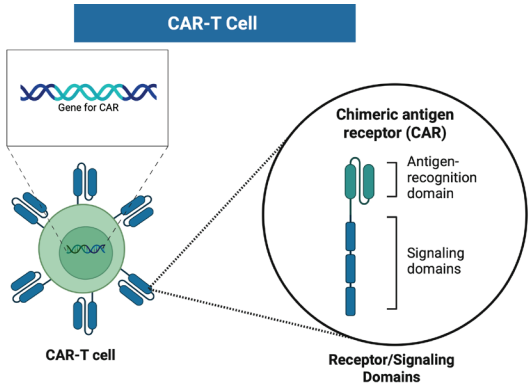


Figure 1: Diagram of a CAR-T cell. Illustrates how the gene for a chimeric antigen receptor (CAR) enables the expression of a synthetic receptor that combines antigen recognition and intracellular signalling domains. (Created with Biorender)

As a personalized treatment, CAR-T cells are engineered lymphocytes, immune cells, that are altered to specifically express certain receptors and target singular cancer cells that display a specific antigen.[3,8] This approach is much more advantageous than regular T cells that rely on a normal major histocompatibility complex (MHC), a group of genes that help the immune system recognize foreign substances, and expression that restrains the immune response.[8] The CAR protein on the external surface of engineered T cells helps recognize cancer antigens, which then activates the CAR-T cell to order the cancer cell to undergo apoptosis (Figure 1).[9] From all this, CAR-T cell therapy has become a very important factor in fighting B-cell malignancies (BCMs) like leukemia.[10] B-cell malignancies are a diverse group of hematologic cancers that have a spectrum that includes aggressive forms like dif-

DOI: 10.36838/v7i10.1

fuse large B-cell lymphoma to indolent variants like follicular lymphoma.[11] These malignancies are characterized by uncontrolled proliferation of malignant B-cells that typically retain surface expression of markers.[11] Malignant B-cells consistently express surface markers, which serve as ideal targets for engineered T-cells.[11,12] For example, acute lymphoblastic leukemia (ALL) expresses CD19, a pan-B-cell maker, making it an ideal target for CAR-T cells.[13] CAR-T cells, therefore, are engineered to target CD19 with high specificity as they bind to the CD19 antigen and trigger cytotoxic responses that eliminate both malignant and normal B cells.[13] However, CAR-T cell therapy faces many obstacles, like patient tolerance and specific antigen selection.[14] Tumour cells that lack specific antigens can impair antigen selectivity of CAR-T cells and can even develop resistance through the downregulation of antigen expression by increasing the activity of immune inhibitory factors from CAR-T cell cytotoxicity.[15,16] In this review, by summarizing the most recent literature on CAR T-cells, the review article will provide the latest research on CAR-T cell therapies in terms of their structure and mechanism, dual antigen targeting advances, optimization strategies for CAR-T cell design, and the limitations and future directions of CAR- T cell therapy research.

Discussion

CAR T Cell Structure and Mechanism:

Chimeric antigen receptor (CAR)-T cell therapy is an innovative immunotherapy and a highly personalized treatment that relies on artificial receptor engineering on T cells to target specific antigens on cancer cells.[17,18] The modular CAR protein is composed of two key components: an external domain for antigen recognition and an internal signalling region for T cell activation. The external region usually consists of a single-chain variable fragment (scFv) derived from antibodies that bind specific antigens on the cancer cell surface through high-affinity binding (Figure 2).[9,19] Unlike traditional T cell receptors (TCRs), this novel design removes the dependency on the MHC, overcoming a major limitation in the internal signalling domain of TCRs. [20,21]

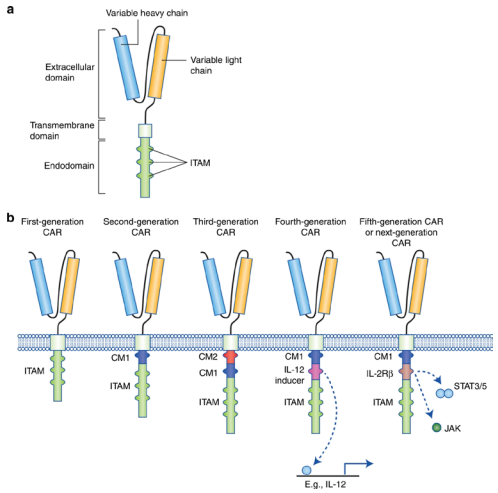


Figure 2: [9] Structure of CAR generations. A. The modular structure of a CAR protein consists of an extracellular domain that includes a single-chain variable fragment (scFv) derived from antibodies, which is responsible for antigen recognition. The transmembrane domain anchors the receptor to the cell membrane and ensures structural stability, and the endodomain contains the intracellular signalling motifs such as immunoreceptor tyrosine-based activation motifs (ITAMs), which mediate downstream T cell activation. B. Demonstrates the progressive development across five generations of CAR-T cells. (Reprints with Permission from Tokarew *et al.*)

Internally, CAR proteins rely on signalling domains for effective T cell activation. The primary activation signal originates from the CD3ξ chain in the intracellular domain, which contains immunoreceptor tyrosine–based activation motifs (ITAMs).[22,23] Additional "boost" signals come from costimulatory domains, like CD28, which help T cells multiply and strengthen their response.[24,25] When fully activated, CAR-T cells release cytotoxic granules that contain perforin and granzymes, which induce apoptosis, self-destruction, in target cells.[26]

Generations of CAR-T cells have had progressive improvements to the overall CAR design. CAR development began with the first-generation CARs. First-generation CARs were comparatively simple in design. It consisted of three main components: an extracellular antigen recognition domain, like a molecular "search radar" to find cancer cells; a transmembrane domain for structural integrity; and an intracellular domain for signalling that activates the T cell (Figure 2b).[9] Single-chain variable fragments (scFvs), which were built for the antigen recognition domain, are engineered antibody fragments composed of a variable heavy chain (VH) and a variable light chain (VL) connected by a flexible peptide linker (Figure 2a).[27] Each scFv is a specialized search tool made by stitching together the two chains from antibodies. The scFvs helped provide high specificity for tumour-associated antigens like CD19 in BCMs like leukemia, for example. This surface protein acts like a cellular "ID card" that is consistently expressed on over 90% of B-cell cancers while being nearly absent from other cell types, minimizing off-target effects.[28,29] Despite the targeted antigen recognition, first-generation CARs relied simply on the CD3ξ signalling domain in the intracellular region, housing the immunoreceptor tyrosine-based activation motifs (ITAMs).[9,30] The ITAMs are responsible for T cell activation upon antigen reception, but were limited due to their signalling capacity. With a lack of co-stimulatory signals to improve capacity, first-generation CARs resulted in poor T cell persistence, reduced cytokine secretion, and limited therapeutic efficacy for first-generation CARs.

Second-generation CARs incorporated co-stimulatory domains such as CD28 to enhance T cell survivability and anti-tumour activity (Figure 2b).[9,31] Alongside CD3ξco-stimulatory signalling domains address many of the limitations of first-generation CARs. CD3ξ chain is the primary signalling domain that contains the ITAM, initiating the costimulatory domain, a secondary signalling domain, obtained from native co-stimulatory molecules like CD28 and 4-1BB, that enhances T cell activation.[32] The extracellular and transmembrane regions remained largely untouched, with their scFvs ensuring targeted binding to tumour antigens like CD19. Most importantly, second-generation CARs became the backbone of FDA-approved CAR-T cell therapies (Table 1).[33]

Third-generation CARs represent a significant advancement in CAR technology through incorporating multiple

costimulatory domains to improve cytokine secretion and overall efficacy (Figure 2b).[9,19] These CARs usually combine two distinct co-stimulatory signals in tandem with CD3ξ. For example, combination domains such as CD28 and 4-1BB are used to provide synergistic signalling: CD28 enhances immediate T cell activation and cytokine release, and 4-1BB promotes T cell survival and memory formation.[34,35] This multi-signal approach has shown improved efficacy, particularly in solid tumours where the TME is highly immunosuppressive. Third-generation CARs, however, are still under investigation.

Fourth-generation CARs, "armoured CARs", are engineered to produce immune-stimulatory cytokines like interleukin-12 (IL-12) to counter immunosuppressive tumour microenvironments (Figure 2b).[9,36] IL-12 is a potent immunomodulatory cytokine that enhances T cell and natural killer cell activity, promotes the recruitment and activation of other immune cells, and reprograms the TME to be more immunogenic to overcome tumour-induced immunosuppression.[37–39] In addition to cytokine secretion, armoured CARs may also express checkpoint inhibitors like PD-1 blockers to prevent T cell exhaustion.[40,41] Enzymes that degrade immunosuppressive factors in the TME, like adenosine or TGF β, are also expressed.[42] These modifications from third-generation CARs make fourth-generation CARs more promising for treating solid tumours, specifically with the degradation of the TME, a major barrier to effective immunotherapy.

Lastly, fifth-generation CARs integrated cytokine receptor signalling to enhance T cell function, representing the cutting edge of CAR-T cell technology (Figure 2b).[9,43] These CARs are designed to mimic the signalling of cytokine receptors like the IL-2 receptor to further enhance T cell activation and persistence. The key innovation in fifth-generation CARs is the inclusion of a cytokine receptor signalling domain alongside the traditional CD3ξ and multiple co-stimulatory domains. This allows the CAR-T cells to respond to endogenous cytokines in the TME, providing an additional layer of activation.

Table 1: FDA Approved CAR-T Cell Therapies.

Name	Brand Name	Target Antigen	Targeted Disease	Reference
Idecabtagene vicleucel	Abecma	anti-BCMA*	Relapsed or Refractory Multiple Myeloma	Munshi NC, Anderson LD Jr, Shah N, *et al. N Engl J Med*. 2021;384(8):705-716. [44]
Lisocabtagene maraleucel	Breyanzi	anti-CD19	Relapsed or Refractory Large B-cell Lymphoma	Sehgal A, Hoda D, Riedell PA, *et al. Lancet Oncol*. 2022;23(8):1066-1077. [45]
Ciltacabtagene autoleucel	Carvykti	anti-BCMA*	Relapsed or Refractory Multiple Myeloma	Berdeja JG, Madduri D, Usmani SZ, *et al. Lancet*. 2021;398(10297):314-324. [46]
Tisagenlecleucel	Kymriah	anti-CD19	B-Cell Lymphoblastic Leukemia	Maude SL, Laetsch TW, Buechner J, *et al. N Engl J Med*. 2018;378(5):439-448. [47]
Brexucabtagene autoleucel	Tecartus	anti-CD19	Mantle Cell Lymphoma	Deshpande A, Wang Y, Munoz J, Jain P. *Drugs Today (Barc)*. 2022;58(6):283-298. [48]
Axicabtagene ciloleucel	Yescarta	anti-CD19	Refractory Large B-Cell Lymphoma	Neelapu SS, Locke FL, Bartlett NL, *et al. N Engl J Med*. 2017;377(26):2531-2544. [49]

*BCMA: B-cell Maturation Antigen

The table highlights six FDA-approved CAR-T cell therapies, each targeting either CD19 or BCMA antigens in various B-cell malignancies. The diversity of targeted diseases underscores the clinical success and expanding role of CAR-T therapy in treating relapsed or refractory hematologic cancers.

CD19-specific CAR T cell therapy, widely used for treating BCMs like leukemia and lymphoma, specifically shows the success of this approach.[12] CD19 is a pan-B-cell marker that ensures selective targeting of malignant cells while sparing most of the healthy tissues.[50] This CD-19 specific approach has demonstrated significant efficacy. In a meta-analysis encompassing 448 patients, 82% had an incidence rate of complete remission (CR).[51] Furthermore, clinical studies report CR rates exceeding 80% in many cases with patients with ALL.[52]

Not only have CD-19-specific strategies seen promising results, but an antibody-drug conjugate targeting CD22, inotuzumab ozogamicin, has shown encouraging results. Antibody-drug conjugates (ADCs), a similar yet different approach to hematologic malignancies, leverage unique molecular characteristics of each patient's tumours through targeting specific antigens like CD19 or CD22.53 ADCs are monoclonal antibodies conjugated to cytotoxic drugs that deliver the drug directly to the tumour and kill the cancer cell internally.[54] Through the selective expression of target antigens on cancer cells, ADCs represent a highly individualized treatment strategy. Inotuzumab ozogamicin in a phase 3 trial demonstrated improved response rates and overall survival compared with conventional methods like chemotherapy in patients with refractory or relapsed B-ALL (Table 2).[55] Blinatumomab is a bispecific T-cell engager (BiTE) antibody that targets the CD3 on T cells and CD19 on B cells. It has significantly improved efficacy in treating relapsed B-cell acute lymphoblastic leukemia (B-ALL) by its mechanism of redirecting T cells to eliminate malignant B cells, leading to improved remission rates.[56] These findings emphasize the efficacy of CD19 and CD22-specific CAR-T cell therapy in treating BCMs, providing substantial benefits as an individualized therapy to patients.

Individualization and Non-Invasive Nature:

Paradigm shifts toward highly individualized cancer treatments have revolutionized health care. CAR-T cell therapy exemplifies this approach. This therapy begins with the collection of the patient's T cells through apheresis, a non-invasive process that separates blood components, minimizing risk to the patient by avoiding the need for invasive surgical intervention.[57] Apheresis collects blood components and returns the rest into the patient, which is essential in minimizing impact (Figure 3). These collected T cells then get genetically engineered to express CARs, designed to recognize tumour antigens on specific tumour cells.[58] Lentiviruses are essential in the addition of CAR genes to the T cells with high efficiency. These viral vectors carry the CAR construct, which includes the intracellular signalling regions and the antigen-recognition domain that are integrated into the host genome ex vivo, ensuring stable expression.[59] Lentiviral vectors are advantageous due to their low immunogenicity and ability to transduce both dividing and nondividing cells.[60,61] The personalized engineering allows a highly specific targeting of malignant cells through uniquely tailoring each CAR-T cell therapy to the patient's cancer profile.[50] With such precision, CAR-T cell therapy reduces the risk of off-target effects, a common issue with conventional treatments like radiation therapy, which often lead to toxicities and significant side effects.

DOI: 10.36838/v7i10.1

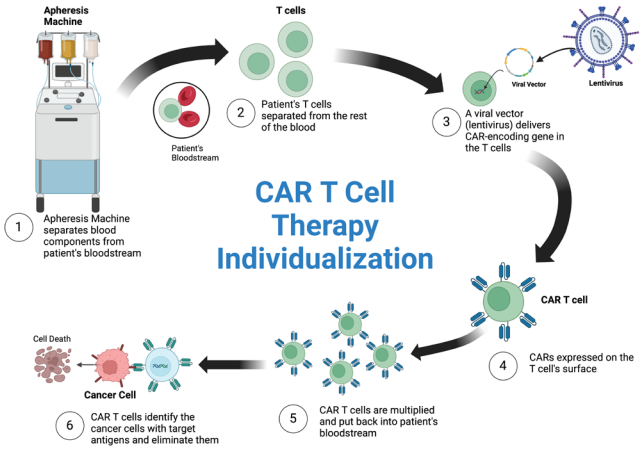


Figure 3: Cycle of the Individualization of CAR-T Cell Therapies. 1. & 2. Isolation of a patient's T cells via apheresis. 3. Genetic modification using a viral vector to express chimeric antigen receptors 4. Receptors are expressed on the T cell's surface 5. Engineered T cells are expanded, multiplied, and reinfused into the patient. 6. CAR T cells eliminate cancer cells with high specificity. This personalized approach enables the targeted recognition and elimination of cancer cells. (Created with Biorender)

The non-invasive nature of CAR-T cell therapy involves only intravenous infusions for cell reinfusion following T cell modification *ex vivo*. Minimizing physical trauma to the patient most often results in shorter recovery times and improved quality of life.[62] Furthermore, CAR-T therapy's ability to minimize residual disease while achieving deep remissions has been highlighted in numerous clinical trials, particularly in hematological malignancies like B-ALL.[63] The individualization of CAR-T cell therapy allows it to address immune evasion and tumour heterogeneity more effectively.[64] CD19-specific CAR-T therapies, for example, target a pan-B-cell marker that is widely expressed in multiple BCMs, which ensures efficacy and reduces the likelihood of tumour escape.[65] Additionally, patients who may not be candidates for aggressive treatments due to age and other complications significantly benefited from CAR-T's non-invasive nature.[66]

Table 2: Clinical Trials on Certain Antibody-Drug Conjugates.

Antibody-Drug Conjugate (ADC)	Target Antigen	Indication	Results	Reference
Blinatumomab	CD19	Relapsed/Refractory B-Cell ALL	34% CR* with full hematological recovery, 44% full, partial, or incomplete hematologic recovery	Kantarjian H, Stein A, Gökbuget N, *et al. N Engl J Med.* 2017;376(9):836-847.[67]
Inotuzumab Ozogamicin	CD22	Relapsed/Refractory B-cell Acute Lymphoblastic Leukemia (B-ALL)	80.7% CR Duration of remission of 4.6 months	Kantarjian HM, DeAngelo DJ, Stelljes M, *et al. N Engl J Med.* 2016;375(8):740-753.[68]
Polatuzumba Vedotin	CD79b	Relapsed/Refractory Diffuse Large B-Cell Lymphoma	77% CR	Tilly H, Morschhauser F, Sehn LH, *et al. N Engl J Med.* 2022;386(4):351-363. [69]
Vadastuximab Talirine	CD33	Acute Myeloid Leukemia (AML)	70% CR and CR with incomplete blood count recovery	Fathi AT, Erba HP, Lancet JE, *et al. Blood.* 2018;132(11):1125-1133. [70]
Loncastuximab Tesirine	CD19	Relapsed/Refractory B-Cell Lymphomas	48.3% ORR** 24.1% CR	Caimi PF, Ai W, Alderuccio JP, *et al. Lancet Oncol.* 2021;22(6):790-800. [71]
Brentuximab Vedotin	CD30	B-cell Lymphomas (Hodgkin Lymphoma)	85% ORR*** 67% CR	Advani RH, Moskowitz AJ, Bartlett NL, *et al. Blood.* 2021;138(6):427-438. [72]

*CR: Complete Response
**ORR: Overall Response Rate
***ORR: Objective Response Rate
The table summarizes the efficacy of various antibody-drug conjugates (ADCs) targeting surface antigens in treating relapsed or refractory hematologic malignancies. The reported high complete response and overall response rates across multiple trials highlight the therapeutic potential of ADCs.

Next-generation CAR-T cell therapies are becoming even more refined as research continues. Through features such as dual antigen targeting, CAR-T cell therapy can enhance patient outcomes and can be featured as the cornerstone of personalized and non-invasive cancer care, potentially inspiring more effective and patient-centred approaches to oncology.

Next-Generation CARs & Optimization Strategies for CAR-T Cell Design

Armoured CARs:

Armoured CAR-T cells, a fourth-generation CAR design, symbolize novel approaches to enhancing the efficacy of CAR-T cell therapy, especially in immunosuppressive tumours.[9] Armoured CARs are engineered to secrete cytokines and other therapeutic molecules like antibodies or enzymes that degrade physical barriers within the tumour microenvironment (TME).[73] The TME is one of the main barriers to CAR-T cell efficacy as it includes extracellular matrix components that hinder CAR-T activation. Through secreting extracellular matrix-degrading enzymes like metalloproteinases, armoured CARs can enhance the persistence and efficient penetration of CAR-T cells in tumours.[74] A study by Suarez *et al.* demonstrates that armoured second-generation CAR-T cells engineered to secrete human anti-programmed death ligand 1 (PD-L1) have effectively blocked PD-1/PD-L1 interaction within the TME of clear cell renal cell carcinoma (ccRCC).[75] Furthermore, a study by Harrasser *et al.* explored the use of armoured CAR-T cells using anti-PD-L1 scFvs for localized delivery within the TME, enhancing the anti-tumour efficacy while minimizing systemic toxicity by blockading the PD-1/PD-L1 pathway.[76] Collectively, these studies indicate the potential of optimizing anti-PD-L1 antibody secretion into next-generation armoured CAR-T cells to enhance their anti-tumour activity by overcoming the immune checkpoint-mediated suppression in the TME.[73,75,76]

Universal CARs:

Driven by the challenge of obtaining autologous T cells from patients for CAR-T cell therapy, researchers have begun to develop universal CAR-T cells. Universal CAR-T cells are allogenic T cells, derived from healthy donors, engineered to target tumour-specific antigens.[77] The key advantage of universal CAR-T cells is the potential of their "universal" use, making CAR-T cell therapy more accessible and scalable.[78] Alternative processes like nanoparticle-mediated in situ manufacturing, for example, can lower the manufacturing procedure, ultimately shortening administration time.[79] This eliminates the logistical challenges and delays associated with autologous CAR-T therapies, like the time required for T-cell engineering. The crucial challenge with allogeneic CAR-T cells is the risk of graft-versus-host disease (GVHD), where the donor T cells attack the patient's tissues. "Off-the-shelf" CAR-T cells, enabled through genetic modification, are engineered to avoid GVHD and prevent immune rejection.[80] This is achieved by removing the TCR in donor T cells, preventing the recognition of patient-specific antigens.[80] Recent trials from Lin *et al*. have shown promising results for universal CAR-T cells, demonstrating that donor-derived CAR-T cells were able to achieve comparable remission rates to autologous CAR-T cells in patients with leukemia.[78]

Advances in Dual Antigen Targeting:

Dual antigen targeting is another key strategy that addresses issues of immune escape and relapse, further benefiting CAR-T therapies in BCMs. Immune escape occurs when cancer cells downregulate or lose their target antigen expression, making CAR T cells effectively useless.[81] This phenomenon is a major player in relapse with patients undergoing CD19 CAR-T cell therapy, relapse due to antigen expression loss.[82,83] BCMs like leukemia and lymphoma often experience relapses due to the loss of CD19 expression, a key target in CAR-T cell therapy.[84] Dual antigen targeting approaches aim to overcome this limitation by developing CAR-T cells that simultaneously target two antigens, such as CD19 and CD37 (Figure 4). This strategy reduces the likelihood of immune escape by ensuring that even if one antigen is lost, the cancer cell would still be susceptible to the CAR-T cell attack. In pre-clinical trials by Scarfò *et al*., anti-CD37 CARs were combined with anti-CD19 CARs to create dual-specific CAR T cells to recognize them alone or in combination, overcoming potential antigen escape.[85] In addition, dual antigen targeting also addresses concerns about heterogeneity within tumour populations. Many BCMs exhibit heterogeneous antigen expression, where many subpopulations of cells lack the primary target antigen.[86] Through dual targeting, dual CAR-T cells ensure coverage across diverse tumour populations, reducing chances of survival or antigen-negative clones.[87] Overall, the durability of the therapeutic response is enhanced even in patients with heterogeneous tumours.

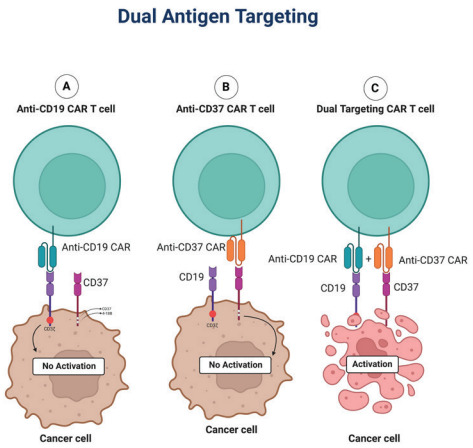


Figure 4: Demonstration of Dual Antigen Targeting CARs. A. Shows the binding of the CD19 antigen to the anti-CD19 CAR. It indicates the activation of only the CD3ξ. B. Shows the binding of the CD37 antigen to the anti-CD37 CAR. It indicates the activation of only the CD37 antigen. C. Shows the binding of both the CD19 and CD37 antigens, resulting in the activation of apoptosis factors in the cancer cell. (Created with Biorender).

Refining dual antigen targeting to improve safety and reduce toxicities is underway. Off-target effects and excessive immune activation can cause adverse effects like cytokine release syndrome (CRS) or neurotoxicity.[88] Researchers are developing strategies like split-signal CARs, requiring simultaneous engagement of both target antigens to activate the CAR-T cells, lowering the risk of off-target activations.[89]

Dual antigen targeting is a promising advancement in CAR-T cell therapy, especially for BCMs prone to antigen loss or heterogeneity. This approach offers a more robust defence against immune escape by targeting multiple antigens, addressing multiple limitations of the current single-antigen CAR-T design. Continued research and trials will further optimize this strategy, potentially making it a significant part of next-generation CAR-T cell therapies.

Controlling CAR-T Cell Activity:

Tightly regulating CAR-T cell activity is crucial to ensuring the safety of this promising approach. Safety switches, engineered mechanisms that enable the deactivation of CAR-T cells in the event of severe adverse effects, are among the many advancements in this field.[90] An example is the inducible suicide gene system, commonly utilizing iCaspase-9, which involves the expression of a caspase protein that gets activated when exposed to a small molecule, which induces apoptosis in the CAR-T cells like AP1903.[91,92] iCaspase-9 has shown rapid means to mitigate life-threatening toxicities like CRS. Clinicians can control CAR-T survivability by activating suicide gene responses to small molecules, which maintains patient safety while preserving the therapeutic benefits of CAR technology.[93]

Regulatable CAR systems allow for the modulation of CAR-T cell activity, which is advantageous in managing toxicity and ensuring that CAR-T therapy is only active when needed.[94] ON-switch CAR, an example of regulatable CAR systems, involves the incorporation of a receptor that can only be activated through the addition of a small molecule.[95] ON-switch CAR designs optimize this strategy through offering

DOI: 10.36838/v7i10.1

an external trigger function, which gives CAR-T flexibility in control. This ability to "turn on" CAR-T cells minimizes off-target effects and reduces the chance of CRS, as the immune cells are only activated by a signal.[96] OFF-switch CARs, relying on molecules to bind to the CAR receptor to deactivate the cells, is a complementary strategy to regulate T cells.[95] This ability to "turn off" CAR-T cells prevents excessive activation and from being overstimulated, which can lead to further toxicity.

Split CAR technology divides the CAR cell into separate components, separating the intracellular signalling domain from the antigen-receptor binding domain.[97] This design requires two molecules to simultaneously bring the components together to activate the CAR-T cells. These designs have improved the treatment of BCMs through targeting CD19.[98] This adds another layer of control to clinicians to manage the availability of the separate components, ensuring the activation only under desired conditions.

Challenges, Limitations, and Future Directions:

Although CAR-T cell therapy has shown success in many clinical trials, there are challenges and limitations that hinder its advancements. So far in the discussion, key limitations are antigen escape, toxicities, and limited efficacies in solid tumours. However, two additional prominent challenges are high cost and safety risks. In patients with severe CRS, it has been estimated to cost up to $500,000 for CAR-T cell therapy.[99] These high prices are constituted of the complex manufacturing process, hospital care, and long-term monitoring, with only specific locations having these therapies.[100] Furthermore, CAR-T cell therapy is linked with severe, life-threatening side effects like cytokine release syndrome.[101] Cytokine release syndrome occurs when massively activated CAR-T cells trigger hyperinflammatory cascades through the excessive release of IL-6, for example.[102] When CAR-T cells become highly activated, they flood the body with inflammatory proteins like IL-6. This "cytokine storm" can damage multiple organs by causing inflammation, seeping blood vessels, and poor oxygen delivery.[103] The heart and blood vessels are often affected, leading to low blood pressure and irregular heartbeats.[103] CRS also frequently impacts the brain, causing confusion, seizures, or even coma in what's called ICANS (immune effector cell-associated neurotoxicity syndrome).[104,105] Treatments include immune-suppressing drugs like tocilizumab and steroids, but severe cases still carry significant risks. Off-target attacks can also lead to severe damage if the target antigen is present in healthy tissues.[106] However, all in all, CAR-T cells' potential is immense and will be vital in future applications. These serious safety concerns, combined with the therapy's high costs, remain major hurdles for CAR-T cell treatment.

Conclusion

In summary, CAR-T cell therapy represents a significant advancement in the treatment of BCMs like leukemia through a highly personalized approach by harnessing the patient's immune system to eliminate cancer cells. Optimization strategies of CAR-T cell design have shown promising success in therapeutic efficacy by reducing relapse and immune escape rates. Advances in the structural engineering of CAR-T cells, like the costimulatory domain incorporation as well as the next-generation CAR design, have improved their ability to overcome the immunosuppressive TME. Universal CAR-T cells advancements have promised better accessibility and minimize adverse effects for a more patient-centred approach. Furthermore, dual antigen targeting strategies have significantly mitigated challenges associated with antigen escape by optimizing the ability to simultaneously recognize multiple tumour-associated antigens. By minimizing relapse, it ensures a more sustained response in leukemia patients. Other safety measures like safety switches and inducible CAR systems have given clinicians more control over CAR-T activity, addressing possible concerns like CRS. Several challenges, like off-target effects and high manufacturing costs, have been logistical hurdles that limit accessibility, which necessitates innovative steps like universal CARs. More research is needed to overcome these hurdles. Future advancements in synthetic biology, like CRISPR and combinatorial therapeutic approaches, may further enhance the benefits of CAR-T cell research. In conclusion, CAR-T cell therapy has demonstrated exponential potential as a transformative immunotherapy for leukemia, with ongoing innovations to improve efficiency. The future of CAR-T cell therapy is bright, paving the way for broader applications in oncology and beyond.

Acknowledgments

I want to thank Dr. Shaye from Stanford University and Dr. Tetz from the University of Michigan for aiding me throughout the completion of this review research paper process.

References

1. Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin*. **2021**, *71* (3), 209–249. https://doi.org/10.3322/caac.21660.
2. Slob, J.; van Houwelingen, T.; Kort, H. S. M. Health Care Work ers' Expectations of the Mercury Advance SMARTcare Solution to Prevent Pressure Injuries: Individual and Focus Group Interview Study. *JMIR Nurs.* **2024**, *7*, e47992. https://doi.org/10.2196/47992.
3. Sun, D.; Shi, X.; Li, S.; Wang, X.; Yang, X.; Wan, M. CAR-T Cell Therapy: A Breakthrough in Traditional Cancer Treatment Strategies (Review). *Mol. Med. Rep.* **2024**, *29* (3), 47. https://doi.org/10.3892/mmr.2024.13171.
4. Chen, H. H. W.; Kuo, M. T. Improving Radiotherapy in Cancer Treatment: Promises and Challenges. *Oncotarget* **2017**, *8* (37), 62742–62758. https://doi.org/10.18632/oncotarget.18409.
5. Paganetti, H.; Beltran, C.; Both, S.; Dong, L.; Flanz, J.; Furutani, K.; Grassberger, C.; Grosshans, D. R.; Knopf, A.-C.; Langendijk, J. A.; Nystrom, H.; Parodi, K.; Raaymakers, B. W.; Richter, C.; Sawakuchi, G. O.; Schippers, M.; Shaitelman, S. F.; Teo, B. K. K.; Unkelbach, J.; Wohlfahrt, P.; Lomax, T. Roadmap: Proton Therapy Physics and Biology. Phys. *Med. Biol.* **2021**, *66* (5). https://doi.org/10.1088/1361-6560/abcd16.
6. Adjiri, A. Identifying and Targeting the Cause of Cancer Is Need ed to Cure Cancer. *Oncol. Ther.* **2016**, *4* (1), 17–33. https://doi.org/10.1007/s40487-015-0015-6.
7. Sterner, R. C.; Sterner, R. M. CAR-T Cell Therapy: Current Limitations and Potential Strategies. *Blood Cancer J.* **2021**, *11* (4), 69. https://doi.org/10.1038/s41408-021-00459-7.
8. De Marco, R. C.; Monzo, H. J.; Ojala, P. M. CAR T Cell Therapy:

A Versatile Living Drug. *Int. J. Mol. Sci.* **2023**, *24* (7), 6300. https://doi.org/10.3390/ijms24076300.
9. Tokarew, N.; Ogonek, J.; Endres, S.; Von Bergwelt-Baildon, M.; Kobold, S. Teaching an Old Dog New Tricks: Next-Generation CAR T Cells. *Br. J. Cancer* **2019,** *120* (1), 26–37. https://doi.org/10.1038/s41416-018-0325-1.
10. Ayala Ceja, M.; Khericha, M.; Harris, C. M.; Puig-Saus, C.; Chen, Y. Y. CAR-T Cell Manufacturing: Major Process Parameters and next-Generation Strategies. *J. Exp. Med.* **2024**, *221* (2), e20230903. https://doi.org/10.1084/jem.20230903.
11. Cuenca, M.; Peperzak, V. Advances and Perspectives in the Treatment of B-Cell Malignancies. *Cancers* **2021**, *13* (9), 2266. https://doi.org/10.3390/cancers13092266.
12. Park, J. H.; Rivière, I.; Gonen, M.; Wang, X.; Sénéchal, B.; Curran, K. J.; Sauter, C.; Wang, Y.; Santomasso, B.; Mead, E.; Roshal, M.; Maslak, P.; Davila, M.; Brentjens, R. J.; Sadelain, M. Long-Term Follow-up of CD19 CAR Therapy in Acute Lymphoblastic Leukemia. *N. Engl. J. Med.* **2018**, *378* (5), 449–459. https://doi.org/10.1056/NEJMoa1709919.
13. Maude, S. L.; Laetsch, T. W.; Buechner, J.; Rives, S.; Boyer, M.; Bittencourt, H.; Bader, P.; Verneris, M. R.; Stefanski, H. E.; Myers, G. D.; Qayed, M.; De Moerloose, B.; Hiramatsu, H.; Schlis, K.; Davis, K. L.; Martin, P. L.; Nemecek, E. R.; Yanik, G. A.; Peters, C.; Baruchel, A.; Boissel, N.; Mechinaud, F.; Balduzzi, A.; Krueger, J.; June, C. H.; Levine, B. L.; Wood, P.; Taran, T.; Leung, M.; Mueller, K. T.; Zhang, Y.; Sen, K.; Lebwohl, D.; Pulsipher, M. A.; Grupp, S. A. Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia. *N. Engl. J. Med.* **2018**, *378* (5), 439–448. https://doi.org/10.1056/NEJMoa1709866.
14. Ahmad, A. CAR-T Cell Therapy. *Int. J. Mol. Sci.* **2020**, *21* (12), 4303. https://doi.org/10.3390/ijms21124303.
15. Majzner, R. G.; Mackall, C. L. Tumor Antigen Escape from CAR T-Cell Therapy. *Cancer Discov.* **2018**, *8* (10), 1219–1226. https://doi.org/10.1158/2159-8290.CD-18-0442.
16. Abreu, T. R.; Fonseca, N. A.; Gonçalves, N.; Moreira, J. N. Current Challenges and Emerging Opportunities of CAR-T Cell Therapies. *J. Control. Release Off. J. Control. Release Soc.* **2020**, *319*, 246–261. https://doi.org/10.1016/j.jconrel.2019.12.047.
17. Porter, D. L.; Levine, B. L.; Kalos, M.; Bagg, A.; June, C. H. Chimeric Antigen Receptor–Modified T Cells in Chronic Lymphoid Leukemia. *N. Engl. J. Med.* **2011**, *365* (8), 725–733. https://doi.org/10.1056/NEJMoa1103849.
18. June, C. H.; O'Connor, R. S.; Kawalekar, O. U.; Ghassemi, S.; Milone, M. C. CAR T Cell Immunotherapy for Human Cancer. *Science* **2018,** *359* (6382), 1361–1365. https://doi.org/10.1126/science.aar6711.
19. Sadelain, M.; Brentjens, R.; Rivière, I. The Basic Principles of Chimeric Antigen Receptor Design. *Cancer Discov.* **2013**, *3* (4), 388–398. https://doi.org/10.1158/2159-8290.CD-12-0548.
20. Jackson, H. J.; Rafiq, S.; Brentjens, R. J. Driving CAR T-Cells Forward. *Nat. Rev. Clin. Oncol.* **2016**, *13* (6), 370–383. https://doi.org/10.1038/nrclinonc.2016.36.
21. Maude, S. L.; Frey, N.; Shaw, P. A.; Aplenc, R.; Barrett, D. M.; Bunin, N. J.; Chew, A.; Gonzalez, V. E.; Zheng, Z.; Lacey, S. F.; Mahnke, Y. D.; Melenhorst, J. J.; Rheingold, S. R.; Shen, A.; Teachey, D. T.; Levine, B. L.; June, C. H.; Porter, D. L.; Grupp, S. A. Chimeric Antigen Receptor T Cells for Sustained Remissions in Leukemia. *N. Engl. J. Med.* **2014**, *371* (16), 1507–1517. https://doi.org/10.1056/NEJMoa1407222.
22. Kalos, M.; Levine, B. L.; Porter, D. L.; Katz, S.; Grupp, S. A.; Bagg, A.; June, C. H. T Cells with Chimeric Antigen Receptors Have Potent Antitumor Effects and Can Establish Memory in Patients with Advanced Leukemia. Sci. *Transl. Med.* **2011**, *3* (95), 95ra73. https://doi.org/10.1126/scitranslmed.3002842.
23. Zhang, X.; Zhu, L.; Zhang, H.; Chen, S.; Xiao, Y. CAR-T Cell Therapy in Hematological Malignancies: Current Opportunities and Challenges. *Front. Immunol.* **2022**, *13*, 927153. https://doi.org/10.3389/fimmu.2022.927153.
24. Kochenderfer, J. N.; Dudley, M. E.; Kassim, S. H.; Somerville, R. P. T.; Carpenter, R. O.; Stetler-Stevenson, M.; Yang, J. C.; Phan, G. Q.; Hughes, M. S.; Sherry, R. M.; Raffeld, M.; Feldman, S.; Lu, L.; Li, Y. F.; Ngo, L. T.; Goy, A.; Feldman, T.; Spaner, D. E.; Wang, M. L.; Chen, C. C.; Kranick, S. M.; Nath, A.; Nathan, D.-A. N.; Morton, K. E.; Toomey, M. A.; Rosenberg, S. A. Chemotherapy-Refractory Diffuse Large B-Cell Lymphoma and Indolent B-Cell Malignancies Can Be Effectively Treated with Autologous T Cells Expressing an Anti-CD19 Chimeric Antigen Receptor. J. Clin. Oncol. Off. *J. Am. Soc. Clin. Oncol.* **2015**, *33* (6), 540–549. https://doi.org/10.1200/JCO.2014.56.2025.
25. Abramson, J. S.; McGree, B.; Noyes, S.; Plummer, S.; Wong, C.; Chen, Y.-B.; Palmer, E.; Albertson, T.; Ferry, J. A.; Arrillaga-Romany, I. C. Anti-CD19 CAR T Cells in CNS Diffuse Large-B-Cell Lymphoma. *N. Engl. J. Med.* **2017**, *377* (8), 783–784. https://doi.org/10.1056/NEJMc1704610.
26. Montalvo, M. J.; Bandey, I. N.; Rezvan, A.; Wu, K.-L.; Saeedi, A.; Kulkarni, R.; Li, Y.; An, X.; Sefat, K. M. S. R.; Varadarajan, N. Decoding the Mechanisms of Chimeric Antigen Receptor (CAR) T Cell-Mediated Killing of Tumors: Insights from Granzyme and Fas Inhibition. *Cell Death Dis.* **2024**, *15* (2), 109. https://doi.org/10.1038/s41419-024-06461-8.
27. Huehls, A. M.; Coupet, T. A.; Sentman, C. L. Bispecific T-Cell Engagers for Cancer Immunotherapy. *Immunol. Cell Biol.* **2015**, *93* (3), 290–296. https://doi.org/10.1038/icb.2014.93.
28. Rousselot, P.; Coudé, M. M.; Gokbuget, N.; Gambacorti Passerini, C.; Hayette, S.; Cayuela, J.-M.; Huguet, F.; Leguay, T.; Chevallier, P.; Salanoubat, C.; Bonmati, C.; Alexis, M.; Hunault, M.; Glaisner, S.; Agape, P.; Berthou, C.; Jourdan, E.; Fernandes, J.; Sutton, L.; Banos, A.; Reman, O.; Lioure, B.; Thomas, X.; Ifrah, N.; Lafage-Pochitaloff, M.; Bornand, A.; Morisset, L.; Robin, V.; Pfeifer, H.; Delannoy, A.; Ribera, J.; Bassan, R.; Delord, M.; Hoelzer, D.; Dombret, H.; Ottmann, O. G.; European Working Group on Adult ALL (EWALL) group. Dasatinib and Low-Intensity Chemotherapy in Elderly Patients with Philadelphia Chromosome-Positive ALL. *Blood* **2016**, *128* (6), 774–782. https://doi.org/10.1182/blood-2016-02-700153.
29. Schuster, S. J.; Bishop, M. R.; Tam, C. S.; Waller, E. K.; Borchmann, P.; McGuirk, J. P.; Jäger, U.; Jaglowski, S.; Andreadis, C.; Westin, J. R.; Fleury, I.; Bachanova, V.; Foley, S. R.; Ho, P. J.; Mielke, S.; Magenau, J. M.; Holte, H.; Pantano, S.; Pacaud, L. B.; Awasthi, R.; Chu, J.; Anak, Ö.; Salles, G.; Maziarz, R. T.; JULIET Investigators. Tisagenlecleucel in Adult Relapsed or Refractory Diffuse Large B-Cell Lymphoma. *N. Engl. J. Med.* **2019**, *380* (1), 45–56. https://doi.org/10.1056/NEJMoa1804980.
30. Smith-Garvin, J. E.; Koretzky, G. A.; Jordan, M. S. T Cell Activation. *Annu. Rev. Immunol.* **2009**, *27*, 591–619. https://doi.org/10.1146/annurev.immunol.021908.132706.
31. Schuster, S. J.; Svoboda, J.; Chong, E. A.; Nasta, S. D.; Mato, A. R.; Anak, Ö.; Brogdon, J. L.; Pruteanu-Malinici, I.; Bhoj, V.; Landsburg, D.; Wasik, M.; Levine, B. L.; Lacey, S. F.; Melenhorst, J. J.; Porter, D. L.; June, C. H. Chimeric Antigen Receptor T Cells in Refractory B-Cell Lymphomas. N. Engl. J. Med. 2017, 377 (26), 2545–2554. https://doi.org/10.1056/NEJMoa1708566.
32. Jayaraman, J.; Mellody, M. P.; Hou, A. J.; Desai, R. P.; Fung, A. W.; Pham, A. H. T.; Chen, Y. Y.; Zhao, W. CAR-T Design: Elements and Their Synergistic Function. *EBioMedicine* **2020**, *58*, 102931. https://doi.org/10.1016/j.ebiom.2020.102931.
33. Chen, Y.-J.; Abila, B.; Mostafa Kamel, Y. CAR-T: What Is Next?

DOI: 10.36838/v7i10.1

Cancers **2023**, *15* (3), 663. https://doi.org/10.3390/cancers15030663.

34. Zhao, Z.; Condomines, M.; van der Stegen, S. J. C.; Perna, F.; Kloss, C. C.; Gunset, G.; Plotkin, J.; Sadelain, M. Structural Design of Engineered Costimulation Determines Tumor Rejection Kinetics and Persistence of CAR T Cells. *Cancer Cell* **2015**, *28* (4), 415–428. https://doi.org/10.1016/j.ccell.2015.09.004.
35. Roselli, E.; Boucher, J. C.; Li, G.; Kotani, H.; Spitler, K.; Reid, K.; Cervantes, E. V.; Bulliard, Y.; Tu, N.; Lee, S. B.; Yu, B.; Locke, F. L.; Davila, M. L. 4-1BB and Optimized CD28 Co-Stimulation Enhances Function of Human Mono-Specific and Bi-Specific Third-Generation CAR T Cells. J. Immunother. *Cancer* **2021**, *9* (10), e003354. https://doi.org/10.1136/jitc-2021-003354.
36. Koneru, M.; Purdon, T. J.; Spriggs, D.; Koneru, S.; Brentjens, R. J. IL-12 Secreting Tumor-Targeted Chimeric Antigen Receptor T Cells Eradicate Ovarian Tumors in Vivo. *Oncoimmunology* **2015**, *4* (3), e994446. https://doi.org/10.4161/2162402X.2014.994446.
37. Pegram, H. J.; Lee, J. C.; Hayman, E. G.; Imperato, G. H.; Ted der, T. F.; Sadelain, M.; Brentjens, R. J. Tumor-Targeted T Cells Modified to Secrete IL-12 Eradicate Systemic Tumors without Need for Prior Conditioning. *Blood* **2012**, *119* (18), 4133–4141. https://doi.org/10.1182/blood-2011-12-400044.
38. Landoni, E.; Woodcock, M. G.; Barragan, G.; Casirati, G.; Cinella, V.; Stucchi, S.; Flick, L. M.; Withers, T. A.; Hudson, H.; Casorati, G.; Dellabona, P.; Genovese, P.; Savoldo, B.; Metelitsa, L. S.; Dotti, G. IL-12 Reprograms CAR-Expressing Natural Killer T Cells to Long-Lived Th1-Polarized Cells with Potent Antitumor Activity. *Nat. Commun.* **2024**, *15* (1), 89. https://doi.org/10.1038/s41467-023-44310-y.
39. Chmielewski, M.; Abken, H. TRUCKs: The Fourth Generation of CARs. *Expert Opin. Biol. Ther.* **2015**, *15* (8), 1145–1154. https://doi.org/10.1517/14712598.2015.1046430.
40. Pang, K.; Shi, Z.-D.; Wei, L.-Y.; Dong, Y.; Ma, Y.-Y.; Wang, W.; Wang, G.-Y.; Cao, M.-Y.; Dong, J.-J.; Chen, Y.-A.; Zhang, P.; Hao, L.; Xu, H.; Pan, D.; Chen, Z.-S.; Han, C.-H. Research Progress of Therapeutic Effects and Drug Resistance of Immunotherapy Based on PD-1/PD-L1 Blockade. Drug Resist. Updat. Rev. Comment. Antimicrob. *Anticancer Chemother.* **2023**, *66*, 100907. https://doi.org/10.1016/j.drup.2022.100907.
41. Rotte, A. Combination of CTLA-4 and PD-1 Blockers for Treatment of Cancer. *J. Exp. Clin. Cancer Res. CR* **2019**, *38* (1), 255. https://doi.org/10.1186/s13046-019-1259-z.
42. Chen, W. TGF-β Regulation of T Cells. *Annu. Rev. Immunol.* **2023**, *41*, 483–512. https://doi.org/10.1146/annurev-immunol-101921-045939.
43. Eyquem, J.; Mansilla-Soto, J.; Giavridis, T.; van der Stegen, S. J. C.; Hamieh, M.; Cunanan, K. M.; Odak, A.; Gönen, M.; Sadelain, M. Targeting a CAR to the TRAC Locus with CRISPR/Cas9 Enhances Tumour Rejection. *Nature* **2017**, *543* (7643), 113–117. https://doi.org/10.1038/nature21405.
44. Munshi, N. C.; Anderson, L. D.; Shah, N.; Madduri, D.; Berdeja, J.; Lonial, S.; Raje, N.; Lin, Y.; Siegel, D.; Oriol, A.; Moreau, P.; Yakoub-Agha, I.; Delforge, M.; Cavo, M.; Einsele, H.; Goldschmidt, H.; Weisel, K.; Rambaldi, A.; Reece, D.; Petrocca, F.; Massaro, M.; Connarn, J. N.; Kaiser, S.; Patel, P.; Huang, L.; Campbell, T. B.; Hege, K.; San-Miguel, J. Idecabtagene Vicleucel in Relapsed and Refractory Multiple Myeloma. *N. Engl. J. Med.* **2021**, *384* (8), 705–716. https://doi.org/10.1056/NEJMoa2024850.
45. Sehgal, A.; Hoda, D.; Riedell, P. A.; Ghosh, N.; Hamadani, M.; Hildebrandt, G. C.; Godwin, J. E.; Reagan, P. M.; Wagner-Johnston, N.; Essell, J.; Nath, R.; Solomon, S. R.; Champion, R.; Licitra, E.; Fanning, S.; Gupta, N.; Dubowy, R.; D'Andrea, A.; Wang, L.; Ogasawara, K.; Thorpe, J.; Gordon, L. I. Lisocabtagene Maraleucel as Second-Line Therapy in Adults with Relapsed or Refractory Large B-Cell Lymphoma Who Were Not Intended for Haematopoietic Stem Cell Transplantation (PILOT): An Open-Label, Phase 2 Study. *Lancet Oncol.* **2022**, *23* (8), 1066–1077. https://doi.org/10.1016/S1470-2045(22)00339-4.
46. Berdeja, J. G.; Madduri, D.; Usmani, S. Z.; Jakubowiak, A.; Agha, M.; Cohen, A. D.; Stewart, A. K.; Hari, P.; Htut, M.; Lesokhin, A.; Deol, A.; Munshi, N. C.; O'Donnell, E.; Avigan, D.; Singh, I.; Zudaire, E.; Yeh, T.-M.; Allred, A. J.; Olyslager, Y.; Banerjee, A.; Jackson, C. C.; Goldberg, J. D.; Schecter, J. M.; Deraedt, W.; Zhuang, S. H.; Infante, J.; Geng, D.; Wu, X.; Carrasco-Alfonso, M. J.; Akram, M.; Hossain, F.; Rizvi, S.; Fan, F.; Lin, Y.; Martin, T.; Jagannath, S. Ciltacabtagene Autoleucel, a B-Cell Maturation Antigen-Directed Chimeric Antigen Receptor T-Cell Therapy in Patients with Relapsed or Refractory Multiple Myeloma (CARTITUDE-1): A Phase 1b/2 Open-Label Study. *Lancet Lond. Engl.* **2021**, *398* (10297), 314–324. https://doi.org/10.1016/S0140-6736(21)00933-8.
47. Maude, S. L.; Laetsch, T. W.; Buechner, J.; Rives, S.; Boyer, M.; Bittencourt, H.; Bader, P.; Verneris, M. R.; Stefanski, H. E.; Myers, G. D.; Qayed, M.; De Moerloose, B.; Hiramatsu, H.; Schlis, K.; Davis, K. L.; Martin, P. L.; Nemecek, E. R.; Yanik, G. A.; Peters, C.; Baruchel, A.; Boissel, N.; Mechinaud, F.; Balduzzi, A.; Krueger, J.; June, C. H.; Levine, B. L.; Wood, P.; Taran, T.; Leung, M.; Mueller, K. T.; Zhang, Y.; Sen, K.; Lebwohl, D.; Pulsipher, M. A.; Grupp, S. A. Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia. *N. Engl. J. Med.* **2018**, *378* (5), 439–448. https://doi.org/10.1056/NEJMoa1709866.
48. Deshpande, A.; Wang, Y.; Munoz, J.; Jain, P. Brexucabtagene Autoleucel: A Breakthrough in the Treatment of Mantle Cell Lymphoma. *Drugs Today* **2022**, *58* (6), 283–298. https://doi.org/10.1358/dot.2022.58.6.3378055.
49. Neelapu, S. S.; Locke, F. L.; Bartlett, N. L.; Lekakis, L. J.; Miklos, D. B.; Jacobson, C. A.; Braunschweig, I.; Oluwole, O. O.; Siddiqi, T.; Lin, Y.; Timmerman, J. M.; Stiff, P. J.; Friedberg, J. W.; Flinn, I. W.; Goy, A.; Hill, B. T.; Smith, M. R.; Deol, A.; Farooq, U.; McSweeney, P.; Munoz, J.; Avivi, I.; Castro, J. E.; Westin, J. R.; Chavez, J. C.; Ghobadi, A.; Komanduri, K. V.; Levy, R.; Jacobsen, E. D.; Witzig, T. E.; Reagan, P.; Bot, A.; Rossi, J.; Navale, L.; Jiang, Y.; Aycock, J.; Elias, M.; Chang, D.; Wiezorek, J.; Go, W. Y. Axicabtagene Ciloleucel CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma. *N. Engl. J. Med.* **2017,** *377* (26), 2531–2544. https://doi.org/10.1056/NEJMoa1707447.
50. Neelapu, S. S.; Tummala, S.; Kebriaei, P.; Wierda, W.; Gutierez, C.; Locke, F. L.; Komanduri, K. V.; Lin, Y.; Jain, N.; Daver, N.; Westin, J.; Gulbis, A. M.; Loghin, M. E.; de Groot, J. F.; Adkins, S.; Davis, S. E.; Rezvani, K.; Hwu, P.; Shpall, E. J. Chimeric Antigen Receptor T-Cell Therapy - Assessment and Management of Toxicities. *Nat. Rev. Clin. Oncol.* **2018**, *15* (1), 47–62. https://doi.org/10.1038/nrclinonc.2017.148.
51. Aamir, S.; Anwar, M. Y.; Khalid, F.; Khan, S. I.; Ali, M. A.; Khattak, Z. E. Systematic Review and Meta-Analysis of CD19-Specific CAR-T Cell Therapy in Relapsed/Refractory Acute Lymphoblastic Leukemia in the Pediatric and Young Adult Population: Safety and Efficacy Outcomes. Clin. *Lymphoma Myeloma Leuk.* **2021**, *21* (4), e334–e347. https://doi.org/10.1016/j.clml.2020.12.010.
52. Cappell, K. M.; Kochenderfer, J. N. Long-Term Outcomes Following CAR T Cell Therapy: What We Know So Far. *Nat. Rev. Clin. Oncol.* **2023**, *20* (6), 359–371. https://doi.org/10.1038/s41571-023-00754-1.
53. Riccardi, F.; Dal Bo, M.; Macor, P.; Toffoli, G. A Comprehensive

Overview on Antibody-Drug Conjugates: From the Conceptualization to Cancer Therapy. *Front. Pharmacol.* **2023**, *14*, 1274088. https://doi.org/10.3389/fphar.2023.1274088.
54. Dumontet, C.; Reichert, J. M.; Senter, P. D.; Lambert, J. M.; Beck, A. Antibody-Drug Conjugates Come of Age in Oncology. Nat. Rev. *Drug Discov.* **2023**, *22* (8), 641–661. https://doi.org/10.1038/s41573-023-00709-2.
55. Kantarjian, H. M.; DeAngelo, D. J.; Stelljes, M.; Martinelli, G.; Liedtke, M.; Stock, W.; Gökbuget, N.; O'Brien, S.; Wang, K.; Wang, T.; Paccagnella, M. L.; Sleight, B.; Vandendries, E.; Advani, A. S. Inotuzumab Ozogamicin versus Standard Therapy for Acute Lymphoblastic Leukemia. *N. Engl. J. Med.* **2016**, *375* (8), 740–753. https://doi.org/10.1056/NEJMoa1509277.
56. Burt, R.; Warcel, D.; Fielding, A. K. Blinatumomab, a Bispecific B-Cell and T-Cell Engaging Antibody, in the Treatment of B-Cell Malignancies. Hum. *Vaccines Immunother.* **2019**, *15* (3), 594–602. https://doi.org/10.1080/21645515.2018.1540828.
57. Bear, A. S.; Fraietta, J. A.; Narayan, V. K.; O'Hara, M.; Haas, N. B. Adoptive Cellular Therapy for Solid Tumors. *Am. Soc. Clin. Oncol. Educ. Book* **2021**, No. 41, 57–65. https://doi.org/10.1200/EDBK_321115.
58. Albinger, N.; Hartmann, J.; Ullrich, E. Current Status and Perspective of CAR-T and CAR-NK Cell Therapy Trials in Germany. *Gene Ther.* **2021**, *28* (9), 513–527. https://doi.org/10.1038/s41434-021-00246-w.
59. Prommersberger, S.; Hudecek, M.; Nerreter, T. Antibody-Based CAR T Cells Produced by Lentiviral Transduction. *Curr. Protoc. Immunol.* **2020**, *128* (1), e93. https://doi.org/10.1002/cpim.93.
60. Labbé, R. P.; Vessillier, S.; Rafiq, Q. A. Lentiviral Vectors for T Cell Engineering: Clinical Applications, Bioprocessing and Future Perspectives. *Viruses* **2021**, *13* (8), 1528. https://doi.org/10.3390/v13081528.
61. Lana, M. G.; Strauss, B. E. Production of Lentivirus for the Es tablishment of CAR-T Cells. *Methods Mol. Biol. Clifton NJ* **2020**, *2086*, 61–67. https://doi.org/10.1007/978-1-0716-0146-4_4.
62. Perthus, A.; Colin, F.; Charton, E.; Anota, A.; Guibert, S. D.; Ysebaert, L.; Daufresne, P.; Lhomme, F.; Manson, G.; Bellec, A.; Bars, L. L.; Chapelle, T. L. D. L.; Houot, R.; Moignet, A. P1687: LIFE AFTER CAR T-CELLS: DO PATIENTS RECOVER A NORMAL LIFE? *HemaSphere* **2023**, *7* (S3), e725933e. https://doi.org/10.1097/01.HS9.0000973620.72593.3e.
63. Maude, S. L.; Frey, N.; Shaw, P. A.; Aplenc, R.; Barrett, D. M.; Bunin, N. J.; Chew, A.; Gonzalez, V. E.; Zheng, Z.; Lacey, S. F.; Mahnke, Y. D.; Melenhorst, J. J.; Rheingold, S. R.; Shen, A.; Teachey, D. T.; Levine, B. L.; June, C. H.; Porter, D. L.; Grupp, S. A. Chimeric Antigen Receptor T Cells for Sustained Remissions in Leukemia. *N. Engl. J. Med.* **2014**, *371* (16), 1507–1517. https://doi.org/10.1056/NEJMoa1407222.
64. Alviano, A. M.; Biondi, M.; Grassenis, E.; Biondi, A.; Serafini, M.; Tettamanti, S. Fully Equipped CARs to Address Tumor Heterogeneity, Enhance Safety, and Improve the Functionality of Cellular Immunotherapies. *Front. Immunol.* **2024**, *15*, 1407992. https://doi.org/10.3389/fimmu.2024.1407992.
65. Milone, M. C.; Xu, J.; Chen, S.-J.; Collins, M. A.; Zhou, J.; Powell, D. J.; Melenhorst, J. J. Engineering Enhanced CAR T-Cells for Improved Cancer Therapy. *Nat. Cancer* **2021**, *2* (8), 780–793. https://doi.org/10.1038/s43018-021-00241-5.
66. Chihara, D.; Liao, L.; Tkacz, J.; Franco, A.; Lewing, B.; Kilgore, K. M.; Nastoupil, L. J.; Chen, L. Real-World Experience of CAR T-Cell Therapy in Older Patients with Relapsed/Refractory Diffuse Large B-Cell Lymphoma. *Blood* **2023**, *142* (12), 1047–1055. https://doi.org/10.1182/blood.2023020197.
67. Kantarjian, H.; Stein, A.; Gökbuget, N.; Fielding, A. K.; Schuh, A. C.; Ribera, J.-M.; Wei, A.; Dombret, H.; Foà, R.; Bassan, R.; Arslan, Ö.; Sanz, M. A.; Bergeron, J.; Demirkan, F.; Lech-Maranda, E.; Rambaldi, A.; Thomas, X.; Horst, H.-A.; Brüggemann, M.; Klapper, W.; Wood, B. L.; Fleishman, A.; Nagorsen, D.; Holland, C.; Zimmerman, Z.; Topp, M. S. Blinatumomab versus Chemotherapy for Advanced Acute Lymphoblastic Leukemia. *N. Engl. J. Med.* **2017**, *376* (9), 836–847. https://doi.org/10.1056/NEJMoa1609783.
68. Kantarjian, H. M.; DeAngelo, D. J.; Stelljes, M.; Martinelli, G.; Liedtke, M.; Stock, W.; Gökbuget, N.; O'Brien, S.; Wang, K.; Wang, T.; Paccagnella, M. L.; Sleight, B.; Vandendries, E.; Advani, A. S. Inotuzumab Ozogamicin versus Standard Therapy for Acute Lymphoblastic Leukemia. *N. Engl. J. Med.* **2016**, *375* (8), 740–753. https://doi.org/10.1056/NEJMoa1509277.
69. Tilly, H.; Morschhauser, F.; Sehn, L. H.; Friedberg, J. W.; Trněný, M.; Sharman, J. P.; Herbaux, C.; Burke, J. M.; Matasar, M.; Rai, S.; Izutsu, K.; Mehta-Shah, N.; Oberic, L.; Chauchet, A.; Jurczak, W.; Song, Y.; Greil, R.; Mykhalska, L.; Bergua-Burgués, J. M.; Cheung, M. C.; Pinto, A.; Shin, H.-J.; Hapgood, G.; Munhoz, E.; Abrisqueta, P.; Gau, J.-P.; Hirata, J.; Jiang, Y.; Yan, M.; Lee, C.; Flowers, C. R.; Salles, G. Polatuzumab Vedotin in Previously Untreated Diffuse Large B-Cell Lymphoma. *N. Engl. J. Med.* **2022**, *386* (4), 351–363. https://doi.org/10.1056/NEJMoa2115304.
70. Fathi, A. T.; Erba, H. P.; Lancet, J. E.; Stein, E. M.; Ravandi, F.; Faderl, S.; Walter, R. B.; Advani, A. S.; DeAngelo, D. J.; Kovacsovics, T. J.; Jillella, A.; Bixby, D.; Levy, M. Y.; O'Meara, M. M.; Ho, P. A.; Voellinger, J.; Stein, A. S. A Phase 1 Trial of Vadastuximab Talirine Combined with Hypomethylating Agents in Patients with CD33-Positive AML. *Blood* **2018**, *132* (11), 1125–1133. https://doi.org/10.1182/blood-2018-03-841171.
71. Caimi, P. F.; Ai, W.; Alderuccio, J. P.; Ardeshna, K. M.; Hamadani, M.; Hess, B.; Kahl, B. S.; Radford, J.; Solh, M.; Stathis, A.; Zinzani, P. L.; Havenith, K.; Feingold, J.; He, S.; Qin, Y.; Ungar, D.; Zhang, X.; Carlo-Stella, C. Loncastuximab Tesirine in Relapsed or Refractory Diffuse Large B-Cell Lymphoma (LOTIS-2): A Multicentre, Open-Label, Single-Arm, Phase 2 Trial. *Lancet Oncol.* **2021**, *22* (6), 790–800. https://doi.org/10.1016/S1470-2045(21)00139-X.
72. Advani, R. H.; Moskowitz, A. J.; Bartlett, N. L.; Vose, J. M.; Ramchandren, R.; Feldman, T. A.; LaCasce, A. S.; Christian, B. A.; Ansell, S. M.; Moskowitz, C. H.; Brown, L.; Zhang, C.; Taft, D.; Ansari, S.; Sacchi, M.; Ho, L.; Herrera, A. F. Brentuximab Vedotin in Combination with Nivolumab in Relapsed or Refractory Hodgkin Lymphoma: 3-Year Study Results. *Blood* **2021**, *138* (6), 427–438. https://doi.org/10.1182/blood.2020009178.
73. Yeku, O. O.; Purdon, T. J.; Koneru, M.; Spriggs, D.; Brentjens, R. J. Armored CAR T Cells Enhance Antitumor Efficacy and Overcome the Tumor Microenvironment. *Sci. Rep.* **2017**, *7* (1), 10541. https://doi.org/10.1038/s41598-017-10940-8.
74. Mehana, E.-S. E.; Khafaga, A. F.; El-Blehi, S. S. The Role of Matrix Metalloproteinases in Osteoarthritis Pathogenesis: An Updated Review. *Life Sci.* **2019**, *234*, 116786. https://doi.org/10.1016/j.lfs.2019.116786.
75. Suarez, E. R.; Chang, D. K.; Sun, J.; Sui, J.; Freeman, G. J.; Signoretti, S.; Zhu, Q.; Marasco, W. A. Chimeric Antigen Receptor T Cells Secreting Anti-PD-L1 Antibodies More Effectively Regress Renal Cell Carcinoma in a Humanized Mouse Model. *Oncotarget* **2016**, *7* (23), 34341–34355. https://doi.org/10.18632/oncotarget.9114.
76. Harrasser, M.; Gohil, S. H.; Lau, H.; Della Peruta, M.; Muczynski, V.; Patel, D.; Miranda, E.; Grigoriadis, K.; Grigoriadis, A.; Granger, D.; Evans, R.; Nathwani, A. C. Inducible Localized Delivery of an Anti-PD-1 scFv Enhances Anti-Tumor Activity of

DOI: 10.36838/v7i10.1

ROR1 CAR-T Cells in TNBC. *Breast Cancer Res.* **2022**, *24* (1), 39. https://doi.org/10.1186/s13058-022-01531-1.
77. Qasim, W. Allogeneic CAR T Cell Therapies for Leukemia. *Am. J. Hematol. 2019*, *94* (S1). https://doi.org/10.1002/ajh.25399.
78. Lin, H.; Cheng, J.; Mu, W.; Zhou, J.; Zhu, L. Advances in Universal CAR-T Cell Therapy. *Front. Immunol.* **2021**, *12*, 744823. https://doi.org/10.3389/fimmu.2021.744823.
79. Fan, M.; Zheng, J.; Huang, Y.; Lu, M.; Shang, Z.; Du, M. Nanoparticle-Mediated Universal CAR-T Therapy. *Int. J. Pharm.* **2024**, *666*, 124779. https://doi.org/10.1016/j.ijpharm.2024.124779.
80. Chen, S.; Van Den Brink, M. R. M. Allogeneic "Off-the-Shelf" CAR T Cells: Challenges and Advances. *Best Pract. Res. Clin. Haematol.* **2024**, *37* (3), 101566. https://doi.org/10.1016/j.beha.2024.101566.
81. Vago, L.; Gojo, I. Immune Escape and Immunotherapy of Acute Myeloid Leukemia. *J. Clin. Invest.* **2020**, *130* (4), 1552–1564. https://doi.org/10.1172/JCI129204.
82. Blass, E.; Ott, P. A. Advances in the Development of Personalized Neoantigen-Based Therapeutic Cancer Vaccines. *Nat. Rev. Clin. Oncol.* **2021**, *18* (4), 215–229. https://doi.org/10.1038/s41571-020-00460-2.
83. Mishra, A.; Maiti, R.; Mohan, P.; Gupta, P. Antigen Loss Following CAR-T Cell Therapy: Mechanisms, Implications, and Potential Solutions. *Eur. J. Haematol.* **2024**, *112* (2), 211–222. https://doi.org/10.1111/ejh.14101.
84. Shalabi, H.; Kraft, I. L.; Wang, H.-W.; Yuan, C. M.; Yates, B.; Delbrook, C.; Zimbelman, J. D.; Giller, R.; Stetler-Stevenson, M.; Jaffe, E. S.; Lee, D. W.; Shern, J. F.; Fry, T. J.; Shah, N. N. Sequential Loss of Tumor Surface Antigens Following Chimeric Antigen Receptor T-Cell Therapies in Diffuse Large B-Cell Lymphoma. *Haematologica* **2018**, *103* (5), e215–e218. https://doi.org/10.3324/haematol.2017.183459.
85. Scarfò, I.; Ormhøj, M.; Frigault, M. J.; Castano, A. P.; Lorrey, S.; Bouffard, A. A.; van Scoyk, A.; Rodig, S. J.; Shay, A. J.; Aster, J. C.; Preffer, F. I.; Weinstock, D. M.; Maus, M. V. Anti-CD37 Chimeric Antigen Receptor T Cells Are Active against B- and T-Cell Lymphomas. *Blood* **2018**, *132* (14), 1495–1506. https://doi.org/10.1182/blood-2018-04-842708.
86. Frigault, M. J.; Maus, M. V. State of the Art in CAR T Cell Therapy for CD19+ B Cell Malignancies. *J. Clin. Invest.* **2020**, *130* (4), 1586–1594. https://doi.org/10.1172/JCI129208.
87. Johnson, D. B.; Nebhan, C. A.; Moslehi, J. J.; Balko, J. M. Immune-Checkpoint Inhibitors: Long-Term Implications of Toxicity. *Nat. Rev. Clin. Oncol.* **2022**, *19* (4), 254–267. https://doi.org/10.1038/s41571-022-00600-w.
88. Hayden, P. J.; Roddie, C.; Bader, P.; Basak, G. W.; Bonig, H.; Bonini, C.; Chabannon, C.; Ciceri, F.; Corbacioglu, S.; Ellard, R.; Sanchez-Guijo, F.; Jäger, U.; Hildebrandt, M.; Hudecek, M.; Kersten, M. J.; Köhl, U.; Kuball, J.; Mielke, S.; Mohty, M.; Murray, J.; Nagler, A.; Rees, J.; Rioufol, C.; Saccardi, R.; Snowden, J. A.; Styczynski, J.; Subklewe, M.; Thieblemont, C.; Topp, M.; Ispizua, Á. U.; Chen, D.; Vrhovac, R.; Gribben, J. G.; Kröger, N.; Einsele, H.; Yakoub-Agha, I. Management of Adults and Children Receiving CAR T-Cell Therapy: 2021 Best Practice Recommendations of the European Society for Blood and Marrow Transplantation (EBMT) and the Joint Accreditation Committee of ISCT and EBMT (JACIE) and the European Haematology Association (EHA). *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **2022**, *33* (3), 259–275. https://doi.org/10.1016/j.annonc.2021.12.003.
89. King-Peoples, T. R.; Posey, A. D. Splitting Signals Drives CARs Further. *Nat. Cancer* **2021**, *2* (9), 873–875. https://doi.org/10.1038/s43018-021-00257-x.
90. Lu, L.; Xie, M.; Yang, B.; Zhao, W.-B.; Cao, J. Enhancing the Safety of CAR-T Cell Therapy: Synthetic Genetic Switch for Spatiotemporal Control. *Sci. Adv.* **2024**, *10* (8), eadj6251. https://doi.org/10.1126/sciadv.adj6251.
91. Guercio, M.; Manni, S.; Boffa, I.; Caruso, S.; Di Cecca, S.; Sinibaldi, M.; Abbaszadeh, Z.; Camera, A.; Ciccone, R.; Polito, V. A.; Ferrandino, F.; Reddel, S.; Catanoso, M. L.; Bocceri, E.; Del Bufalo, F.; Algeri, M.; De Angelis, B.; Quintarelli, C.; Locatelli, F. Inclusion of the Inducible Caspase 9 Suicide Gene in CAR Construct Increases Safety of CAR.CD19 T Cell Therapy in B-Cell Malignancies. *Front. Immunol.* **2021**, *12*, 755639. https://doi.org/10.3389/fimmu.2021.755639.
92. Zhou, X.; Di Stasi, A.; Brenner, M. K. iCaspase 9 Suicide Gene System. Methods Mol. Biol. *Clifton NJ* **2015**, *1317*, 87–105. https://doi.org/10.1007/978-1-4939-2727-2_6.
93. Stock, S.; Klüver, A.-K.; Fertig, L.; Menkhoff, V. D.; Subklewe, M.; Endres, S.; Kobold, S. Mechanisms and Strategies for Safe Chimeric Antigen Receptor T-Cell Activity Control. *Int. J. Cancer* **2023**, *153* (10), 1706–1725. https://doi.org/10.1002/ijc.34635.
94. Labanieh, L.; Majzner, R. G.; Klysz, D.; Sotillo, E.; Fisher, C. J.; Vilches-Moure, J. G.; Pacheco, K. Z. B.; Malipatlolla, M.; Xu, P.; Hui, J. H.; Murty, T.; Theruvath, J.; Mehta, N.; Yamada-Hunter, S. A.; Weber, E. W.; Heitzeneder, S.; Parker, K. R.; Satpathy, A. T.; Chang, H. Y.; Lin, M. Z.; Cochran, J. R.; Mackall, C. L. Enhanced Safety and Efficacy of Protease-Regulated CAR-T Cell Receptors. *Cell* **2022**, *185* (10), 1745-1763.e22. https://doi.org/10.1016/j.cell.2022.03.041.
95. Zhang, J.; Du, B.; Liu, M. Let's Turn the CAR-T Cells ON and OFF Precisely. *Cancer Cell* **2022**, *40* (11), 1264–1266. https://doi.org/10.1016/j.ccell.2022.10.019.
96. Jan, M.; Scarfò, I.; Larson, R. C.; Walker, A.; Schmidts, A.; Guirguis, A. A.; Gasser, J. A.; Słabicki, M.; Bouffard, A. A.; Castano, A. P.; Kann, M. C.; Cabral, M. L.; Tepper, A.; Grinshpun, D. E.; Sperling, A. S.; Kyung, T.; Sievers, Q. L.; Birnbaum, M. E.; Maus, M. V.; Ebert, B. L. Reversible ON- and OFF-Switch Chimeric Antigen Receptors Controlled by Lenalidomide. *Sci. Transl. Med.* **2021**, *13* (575), eabb6295. https://doi.org/10.1126/scitranslmed.abb6295.
97. Li, S.; Shi, L.; Zhao, L.; Guo, Q.; Li, J.; Liu, Z.; Guo, Z.; Cao, Y. J. Split-Design Approach Enhances the Therapeutic Efficacy of Ligand-Based CAR-T Cells against Multiple B-Cell Malignancies. *Nat. Commun.* **2024**, *15* (1), 9751. https://doi.org/10.1038/s41467-024-54150-z.
98. He, X.; Feng, Z.; Ma, J.; Ling, S.; Cao, Y.; Gurung, B.; Wu, Y.; Katona, B. W.; O'Dwyer, K. P.; Siegel, D. L.; June, C. H.; Hua, X. Bispecific and Split CAR T Cells Targeting CD13 and TIM3 Eradicate Acute Myeloid Leukemia. *Blood* **2020**, *135* (10), 713–723. https://doi.org/10.1182/blood.2019002779.
99. Choi, G.; Shin, G.; Bae, S. Price and Prejudice? The Value of Chimeric Antigen Receptor (CAR) T-Cell Therapy. Int. J. Environ. Res. *Public. Health* **2022**, *19* (19), 12366. https://doi.org/10.3390/ijerph191912366.
100. Cliff, E. R. S.; Kelkar, A. H.; Russler-Germain, D. A.; Tessema, F. A.; Raymakers, A. J. N.; Feldman, W. B.; Kesselheim, A. S. High Cost of Chimeric Antigen Receptor T-Cells: Challenges and Solutions. *Am. Soc. Clin. Oncol. Educ. Book* **2023**, No. 43, e397912. https://doi.org/10.1200/EDBK_397912.
101. Frey, N.; Porter, D. Cytokine Release Syndrome with Chimeric Antigen Receptor T Cell Therapy. Biol. Blood Marrow Transplant. J. Am. Soc. *Blood Marrow Transplant.* **2019**, *25* (4), e123–e127. https://doi.org/10.1016/j.bbmt.2018.12.756.
102. Chen, P.; Tang, Y.; He, W.; Yang, R.; Lan, Z.; Chen, R.; Zhang, P. Potential Pathophysiological Mechanisms Underlying Multiple Organ Dysfunction in Cytokine Release

Syndrome. *Mediators Inflamm.* **2022**, *2022*, 7137900. https://doi.org/10.1155/2022/7137900.
103. Müller, F.; Taubmann, J.; Bucci, L.; Wilhelm, A.; Bergmann, C.; Völkl, S.; Aigner, M.; Rothe, T.; Minopoulou, I.; Tur, C.; Knitza, J.; Kharboutli, S.; Kretschmann, S.; Vasova, I.; Spoerl, S.; Reimann, H.; Munoz, L.; Gerlach, R. G.; Schäfer, S.; Grieshaber-Bouyer, R.; Korganow, A.-S.; Farge-Bancel, D.; Mougiakakos, D.; Bozec, A.; Winkler, T.; Krönke, G.; Mackensen, A.; Schett, G. CD19 CAR T-Cell Therapy in Autoimmune Disease - A Case Series with Follow-Up. *N. Engl. J. Med.* **2024**, *390* (8), 687–700. https://doi.org/10.1056/NEJMoa2308917.
104. Freyer, C. W.; Porter, D. L. Cytokine Release Syndrome and Neurotoxicity Following CAR T-Cell Therapy for Hematologic Malignancies. *J. Allergy Clin. Immunol.* **2020**, *146* (5), 940–948. https://doi.org/10.1016/j.jaci.2020.07.025.
105. Gazeau, N.; Liang, E. C.; Wu, Q. V.; Voutsinas, J. M.; Barba, P.; Iacoboni, G.; Kwon, M.; Ortega, J. L. R.; López-Corral, L.; Hernani, R.; Ortiz-Maldonado, V.; Martínez-Cibrian, N.; Martinez, A. P.; Maziarz, R. T.; Williamson, S.; Nemecek, E. R.; Shadman, M.; Cowan, A. J.; Green, D. J.; Kimble, E.; Hirayama, A. V.; Maloney, D. G.; Turtle, C. J.; Gauthier, J. Anakinra for Refractory Cytokine Release Syndrome or Immune Effector Cell-Associated Neurotoxicity Syndrome after Chimeric Antigen Receptor T Cell Therapy. *Transplant. Cell. Ther.* **2023**, *29* (7), 430–437. https://doi.org/10.1016/j.jtct.2023.04.001.
106. Flugel, C. L.; Majzner, R. G.; Krenciute, G.; Dotti, G.; Riddell, S. R.; Wagner, D. L.; Abou-El-Enein, M. Overcoming On-Target, off-Tumour Toxicity of CAR T Cell Therapy for Solid Tumours. *Nat. Rev. Clin. Oncol.* **2023**, *20* (1), 49–62. https://doi.org/10.1038/s41571-022-00704-3.

Author

Alex Zhang is a Senior at Fraser Heights Secondary School. He hopes to pursue his passion in biomedicine and biotechnology to follow his dreams in medicine.

DOI: 10.36838/v7i10.1

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

ESG Score Prediction with Financial Metrics

Aravindkrishna Arivudainambi

Lynbrook High School, 1280 Johnson Ave, San Jose, California, 95129, U.S.A.; aravindkrishna.nambi@gmail.com

ABSTRACT: This study investigates the relationship between various financial metrics and companies' Environmental, Social, and Governance (ESG) scores. Based on a comprehensive analysis of financial data, the research aims to develop a machine-learning model capable of predicting ESG scores. Leveraging datasets from Kaggle and Yahoo Finance, we extracted relevant information, including cash flow statements, income statements, balance sheets, and company-specific attributes such as sector, location, and employee count. The methodology involved data preprocessing, feature engineering, and evaluating multiple regression models, including multilayer perceptron regressor, linear regression, lasso regression, ridge regression, and elastic net regression. The results demonstrate that the Multilayer Perceptron Regressor outperformed the other models, achieving a mean squared error of 0.145 and an R-squared value of 0.989. This indicates its superior ability to capture the complex relationships between financial metrics and ESG scores. The study highlights the significance of financial data in assessing a company's ESG performance and provides insights into the most influential factors contributing to ESG scores. Additionally, it underscores the potential of machine learning techniques in predicting sustainability and ethical practices based on quantitative financial information. The findings have practical implications for investors, analysts, and companies seeking to align their financial strategies with environmental, social, and governance considerations.

KEYWORDS: Robotics and Intelligence Machines, Machine Learning, Finance, Environmental Science, Sustainability.

Introduction

Environmental, Social, and Governance (ESG) factors have gained significant prominence in the investment landscape in recent years. Investors, stakeholders, and regulatory bodies increasingly seek to align their financial decisions with sustainable and ethical practices. As a result, there is a growing demand for accurate assessment and prediction of companies' ESG performance.[1] This study aims to leverage the power of machine learning techniques to develop a predictive model capable of estimating ESG scores based on comprehensive financial data.

Historically, ESG evaluations have relied heavily on qualitative assessments and subjective criteria.[2] However, this approach may overlook the intricate relationships between quantitative financial metrics and a company's commitment to sustainability and ethical governance.[3] By harnessing the wealth of information contained within financial statements, such as cash flow statements, income statements, and balance sheets, this research endeavors to uncover the quantitative underpinnings of ESG performance.

The methodology employed in this study involves data acquisition from reputable sources, meticulous feature engineering, and the evaluation of multiple regression models, including advanced techniques like the Multilayer Perceptron Regression. The overarching goal is to identify the most influential financial factors contributing to a company's ESG score and develop a robust predictive model that can assist investors, analysts, and companies in aligning their financial strategies with environmental, social, and governance considerations.

The study found that financial metrics are important in contributing to a company's ESG score. Specifically, the MLP regressor is the model with the most prominent result due to its more complex structure, which can detect more minute patterns within the data. Other regressors used, including linear, Lasso, Ridge, and Elastic Net, provided wretched results with high error and low R-squared values.

Literature Review

One of the first pieces of literature I looked at was the "Empirical Study of ESG Score Prediction through Machine Learning—A Case of Non-Financial Companies in Taiwan" by Hsio-Yi Lin and Bin-Wei Hsu. They use machine learning models to predict ESG scores for Taiwan pre-COVID times. Their results are about the performances of different models, and the models that came to the top were RF, ELM, SVM, and XGBoost. The article was created before 2019 and uses Taiwanese datasets, so the results might not completely match our results, but should align closely. They split each category of importance into different features, e.g., the ESG score is divided into environmental, social, and governance scores. Likewise, the financial metric they used was long-term capital adequacy ratio, current ratio, quick ratio, fixed assets turnover ratio, return on operating assets ROA before tax, and operating profit to paid-up capital ratio. The document also splits governance indicators into multiple categories, including but not limited to stock earnings deviation, earnings per share deviation ratio, and total shares. Their results showed significantly high R-squared values, greater than 0.98, for most of their test data amounts.[4] The metrics they used are closely related to what we are using.

DOI: 10.36838/v7i10.12

The following article I looked at was "Capital Structure and Speed of Adjustment: The Impact of Environmental, Social, and Governance (ESG) Performance" by Yusuf Babatunde Adenye and Ines Kammoun. This study tries to see how a company's efforts in being environmentally, socially, and ethically responsible affect how much debt it takes on and how quickly it adjusts its debt to meet its needs. Their results showed that companies with higher ESG must borrow more but are perceived as less reliant on debt by the market. Those excelling in environmental aspects adjust their debt levels faster to meet their finances.[5] This made me think about debt and its impact on the ESG score, and I realized that it was an important metric to consider when calculating ESG scores based on financial metrics.

The final article I believed impacted the experiment was an economic article for predicting ESG Scores (https://www.economy.com/products/esg/esg-score-predictor). This article showed how they can accurately predict ESG scores using sector and location data. This was my first thought of using sector and location, but then I realized that it wouldn't be as useful because I only have fewer than 500 data points, and they are all in a relatively nearby area. This data covers many countries and jurisdictions, so they have more overall data. This would result in a more accurate model that predicts ESG scores. The article does not provide much information about the training process other than the amount of data they used.[6]

Methods

To achieve this goal, data was acquired from Kaggle and Yahoo Finance.[7,8] The Kaggle dataset provided information about S&P 500 companies and their respective ESG ratings.[7] The dataset also provides the sector and industry of a company, which has a relationship with the ESG scores. The dataset also includes the address, full-time employees, controversy score, and controversy level. A business's location affects a company's environmental score and seems to be an important feature of the model. Other information, including full-time employees, controversy score, and controversy level, could provide insight into the ESG score. We extracted each company's cash flow, balance sheet, and income statement from the Yahoo Finance dataset.[7,9] This data was necessary because ESG is commonly used to get investors and capital into a company. Using this knowledge, there can be a correlation between data from cash flow, balance sheet, income statement, and ESG Scores. In the world of corporate sustainability and responsible investing, evaluating Environmental, Social, and Governance (ESG) scores is crucial.[10] Different industries have unique criteria for assessing performance within the ESG framework. For example, energy companies are closely examined for carbon emissions, while the technology sector focuses on data privacy and security for the social aspect of ESG scores. However, analyzing specific sectors, like the S&P 500, can be challenging due to the limited number of companies. Striking a balance between sector-specific insights and broader trends is necessary to thoroughly assess a company's ESG performance, highlighting the complex relationship between industry dynamics and sustainability metrics.

Geographical factors significantly influence a company's ESG position, with location data playing a crucial role.[11] The region of operation can impact environmental, social, and governance practices. For example, companies in areas with strict environmental regulations may have lower carbon footprints, positively affecting their ESG scores in the environmental category. Social practices are influenced by cultural and societal norms in the company's location, and governance standards vary based on regional regulatory frameworks. Integrating location data into ESG assessments enhances understanding by considering contextual factors that shape a company's sustainability and ethical practices.

Full-time employees, controversy score, and controversy level may not provide meaningful insights into the ESG score as they often lack direct correlations with the environmental, social, and governance factors that comprise the ESG framework. While important operationally, the number of full-time employees may not inherently reflect a company's sustainability practices. Similarly, while indicating potential ethical concerns, the controversy score and controversy level may not align with the criteria evaluated within the ESG framework, making them less directly relevant to the comprehensive assessment of a company's ESG performance. Therefore, these factors may offer limited value when evaluating a company's sustainability and ethical standing within the ESG context.

Examining the Cash Flow statement on Yahoo Finance is crucial for calculating a company's ESG score. This financial metric provides insights into environmental, social, and governance (ESG) aspects.5 Positive cash flow not only supports environmentally sustainable initiatives (Environmental Score) but also allows investments in social programs (Social Score). Efficient cash flow management indicates strong governance practices, enhancing transparency and financial stewardship (Governance Score). Analyzing the Cash Flow statement is essential for comprehensively understanding a company's commitment to sustainability, social responsibility, and effective governance within the ESG framework.

The income statement is crucial in establishing a link between a company's financial performance and ESG score. Providing a comprehensive overview of revenues, expenses, and profits, the income statement is instrumental in evaluating a company's environmental, social, and governance (ESG) aspects. The profits and expenditures disclosed in this statement allow the assessment of a company's commitment to environmentally sustainable practices, which influence the environmental score. Additionally, the income statement provides insights into financial allocations for social responsibility initiatives, impacting the Social Score. As the income statement reflects, transparent financial reporting and responsible resource allocation contribute to strong governance practices, positively affecting the Governance Score. A thorough examination of the income statement facilitates a holistic understanding of a company's financial health and alignment with ESG principles, guiding investors and analysts in assessing its sustainability and ethical practices.

Completing the financial triad, the balance sheet is crucial for comprehensively understanding a company's ESG stand-

ing.[12] Detailing assets, liabilities, and equity provides key insights into a company's financial health and its implications for environmental, social, and governance (ESG) considerations. The asset side may reveal investments in sustainable resources, contributing to the Environmental Score. Liabilities and equity illuminate a company's financial obligations and ownership structure, influencing social responsibility and governance practices, respectively. A strong and transparent balance sheet reflects sound financial management, positively impacting the Governance Score. Thoroughly analyzing the balance sheet is integral for investors and analysts, as it assesses a company's overall financial stability and alignment with ESG principles and completes the tripartite evaluation of sustainability and ethical practices.

To achieve the best use of all our data, we extracted multiple columns from all three of the financial triad. The first item we want is the return on equity. We will use assets, liabilities, and net income to determine the shareholders' equity. We want to know how much of the company's worth is made by the investors of the capital. The next item we look for is the Current Ratio. The current ratio is useful for finding a company's ability to cover its liabilities with its assets in the short term.[13] The next item we look for is earnings per share. This value shows how many shares there are to purchase in the market. The more you must split your ownership with others.[14] The next item we concatenated is the net profit margin. This shows how much profit you earn from each product. The final items we used are free cash flow, which checks if there is money to fund accidental incidents, and operating cash flow, which shows how much cash we have aside for operating.[15,16] The table below shows how we computed all these values and how they are defined (Table 1).

Table 1: Derived financial metrics and definitions. The table shows the importance of each financial metric.

	Derivation	Description
Return on Equity (ROE)	$ROE = \frac{Net\ Income}{Total\ Assets - Total\ Liabilities}$	Used to find the shareholder's equity. Will provide how much of the company's total worth is made of the investor's capital
Current Ratio (CR)	$CR = \frac{Current\ Asset}{Current\ Liability}$	Finding the company's ability to cover its liabilities with its assets in the short term
Long-Term Debt to Net Income (LTDNI)	$LTDNI = \frac{Long\ Term\ Debt}{Net\ Income}$	This tells how many years it will take for a company to pay back its long-term debt with its income. A more sufficient company should pay it back faster.
Earnings per Share (EPA)	$EPA = \frac{Net\ Income}{Shares\ Issued}$	How many shares are there to purchase in the market, the more you must split your ownership with others
Net Profit Margin (NPM)	$NPM = \frac{Net\ Income}{Total\ Revenue}$	How much does the company earn from each product
Free Cash Flow	N/A	Check if there is money to fund accidental incidents
Operating Cash Flow	N/A	How much cash does the company have set aside for operating
Net Earnings	N/A	Profit the company makes

Next was preprocessing the data. To achieve this goal, we first removed all the missing data, which removed six rows from the data. We didn't use imputation because only a small amount of data needed to be removed; imputation seemed unnecessary. We also did not need to remove any outliers because we were taking data from the last year. Because of this, there is no outlier comparison to conduct.

Standardizing the data with sklearn's StandardScaler, where each data point is transformed by subtracting the mean and dividing by the standard deviation (represented as z=xi-μ), is an essential preprocessing step. This step is important because it makes all the data on the same scale. Standardizing facilitates accurate model training and improves the model's adaptability to varying datasets, fostering a more resilient and dependable predictive performance. Additionally, it enhances the model's ability to handle outliers and ensures more effective convergence during the training process.

Next, we removed the unimportant features of the dataset. We only focused on the data we derived and deemed all the other information unnecessary. The main idea behind this is that all the data we thought necessary, as discussed in Table 1, have incorporated all the required data between each financial triad. One of the main reasons for this is that many columns in the other parts of the data were missing, so removing this would provide very small amounts of data to train with. We cherry-picked the most important features based on two qualities.

Additionally, the feature importance of the given variables was indicative of the factors most responsible for the variance in ESG scores.

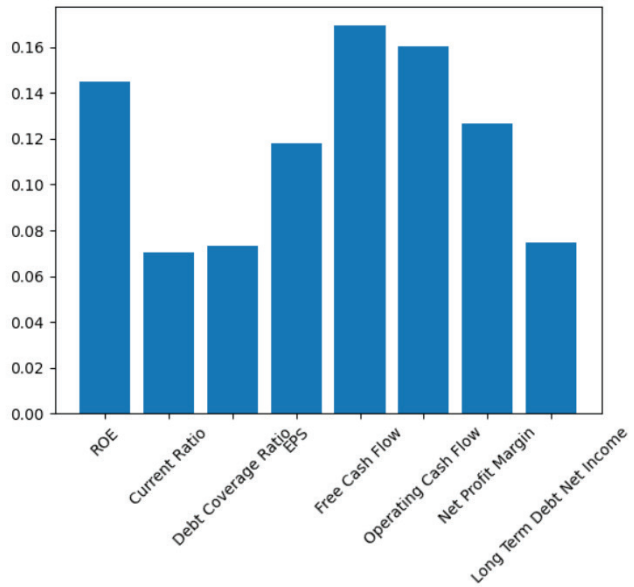


Figure 1: Financial metric feature importance bar chart. The most important feature is free cash flow. All features are significant to the model because each feature's importance is above 0.05.

If the feature importance is above 0.05, the feature is valuable in the training process. This graph shows the different features and their corresponding importance, and all the values are above 0.05, showing their contribution to the model. The highest of these is the free cash flow, which shows its highest importance. We first train a model to calculate these scores and then list the weights for each feature. The importance of the weights is listed in the graph above. The data we believe will

DOI: 10.36838/v7i10.12

provide optimal results includes ROE, Current Ratio, Debt Coverage Ratio, EPS, Free Cash Flow, Operating Cash Flow, Net Profit Margin, Current Ratio, Long Term Debt Net Income, Location, and Sector. After further analysis, we decided to remove location from the data because all the companies are inside the United States and can skew the data more than help it. We also removed the sector from the dataset because the range of the sector was too broad and would be too diluted in the dataset to provide less gain for the model to pick up on.

The next step to train the model was splitting the dataset into training and testing piles. We used an 80:20 ratio for training and testing. We tested multiple regression-based models, including MLP Regressor, Linear Regression, Lasso Regression, Ridge Regression, and Elastic Net Regression.

The Multilayer Perceptron Regressor is a neural network model that excels in capturing intricate nonlinear relationships within data. In our situation, where we are dealing with a diverse set of financial features and aiming to predict complex ESG scores, the MLP Regressor is well-suited. Its ability to automatically learn and adapt to patterns in large datasets aligns with the complexity of factors influencing ESG performance. However, caution is required as MLPs are prone to overfitting, especially without careful hyperparameter tuning. The computational expense is another consideration, but given the significance of the task, the MLP Regressor offers a powerful tool for modeling the intricate connections between financial metrics and ESG scores. However, unlike many of the models discussed below, the MLP regressor can fit a nonlinear relationship between the explanatory and response variables.[17]

Linear Regression is a fundamental and interpretable model that assumes a linear relationship between input features and the target variable. While this model might seem simplistic, its straightforward nature makes it an important baseline for comparison. The linearity may not fully capture the complex connections present in our data. This is calculated using the formula $Y_i = f(X_i,\beta) + e_i$, where Y_i, which is the dependent variable, f is the function, X_i the independent variables, β are the unknown parameters and e_i is the error terms.[18]

Lasso Regression, a variant of linear regression, introduces regularization by adding a penalty term based on the absolute values of the coefficients. This encourages sparsity, effectively performing feature selection. In our case, where we have diverse financial features, Lasso Regression becomes valuable for identifying the most influential variables for predicting ESG scores. It aids in streamlining the model by focusing on the most relevant metrics and preventing overfitting. However, careful tuning of the regularization parameter is crucial to balance the feature selection and maintain model performance.[19,20]

Ridge Regression is another variant of linear regression, but with regularization based on the squared values of the coefficients. It effectively handles multicollinearity, making it suitable for situations where financial metrics may be correlated. In our context, when dealing with a triad of financial data – cash flow, income statement, and balance sheet – Ridge Regression becomes a valuable tool. It helps prevent overfitting and ensures that the model generalizes well to the complex relationships between financial health and ESG scores.[21]

Elastic Net Regression combines the strengths of L1 (Lasso) and L2 (Ridge) regularization, offering a balanced approach. It is beneficial when dealing with datasets that may have correlated features. In our scenario, where we explore the relationship between various financial metrics and ESG scores, Elastic Net Regression becomes relevant. Simultaneously, addressing feature selection and multicollinearity provides a robust framework for modeling the intricate factors contributing to ESG performance. However, the challenge lies in tuning two hyperparameters, making them more complex than individual Lasso or Ridge models. Careful optimization is required to strike the right balance and harness the strengths of both regularization techniques. The table below summarizes all the models and their positives and negatives (Table 2).[22,23]

Table 2: Model trade-off analysis based on machine learning theory.

	Advantages	Disadvantages
MLP Regressor	Capable of handling complex patterns	Prone to overfitting, Computational expense, requires careful hyperparameter tuning
Linear Regression	Easy to interpret and understand	Limited to linear patterns
Lasso Regression	Performs feature selection	May struggle with highly correlated features
Ridge Regression	Effective in handling multicollinearity	May not detect patterns that are not easily visible
Elastic Net Regression	Balances feature selection and multicollinearity	May not detect patterns that are not easily visible, More complex than Lasso or Ridge alone

We then trained the models and checked for two different values to see their accuracy. The first value we looked for is the mean squared error. The mean squared error (MSE) measures how close a regression line is to a set of data points. This means the higher the MSE, the less accurate the model is, and the lower the MSE, the more accurate the model is. The subsequent measurement we used to test the accuracy of our model is the R-squared, which we used for the line of best fit at the start of this section. It is a statistical measure that provides information about the goodness of fit of a regression model. The closer the model is to 1.0, the better the model performs.

Results and Discussion

We first measured the individual correlations between each item and its correlations with their ESG score. The results for some scatterplots and their corresponding R-squared measurements are shown.

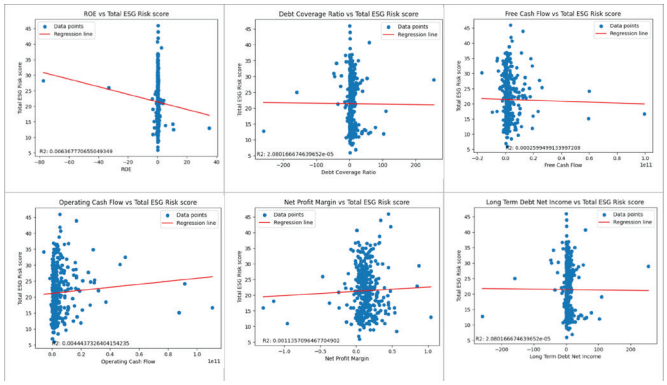


Figure 2: Scatterplot of Financial Metrics vs ESG Score. No apparent correlation is seen between the values.

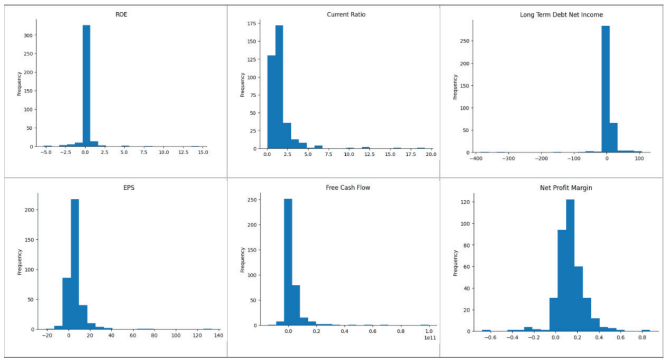


Figure 3: Distribution of financial metrics. Most are normally distributed; the current ratio and EPS are skewed right.

Most graphs, including the current ratio and free cash flow, are skewed to the left. ROE looks evenly distributed, with most items close to 0.0 (Figure 3). The Net Profit Margin is also normally distributed, with most items close to 0.1. The current ratio is skewed to the right. If the items are skewed to the right, they have more power, like for EPS. Skewed left means that the values are closer to 0.

Using all this knowledge from the data histograms and scatter plots, we trained the data with an MLP Regressor, Linear Regression, Lasso Regression, Ridge Regression, and Elastic Net Regression. The results in Table 3 show corresponding R-squared and MSE values and indicate that the MLP regressor is the clear best model. The subsequent section analyzes why the models performed the way they did.

Table 3: Mean Squared Error and R-squared value attained by each model.

	Mean Squared Error (MSE)	R^2
MLP Regressor	0.145	0.989
Linear Regression	18.394	-0.419
Lasso Regression	22.217	-0.714
Ridge Regression	18.475	-0.426
Elastic Net Regression	18.475	-0.426

The R-squared is negative in some cases because this signifies that the model does not follow the trend of the data.[24] Another way to put this is that the selected model is worse than choosing a horizontal line for this data. This shows how badly the regressor fits the data. The Mean Squared Error (MSE) and R-squared (R-squared) values provide insights into the performance of different regression models. The MLP Regressor achieved a low MSE of 0.145 and a high R-squared of 0.989, indicating its better predictive accuracy and ability to explain variance in the data.

In contrast, Linear Regression, Lasso Regression, Ridge Regression, and Elastic Net Regression all exhibited higher MSE values and negative R-squared scores, suggesting poor model fit and limited ex-planatory power. These traditional linear models may struggle to capture the underlying patterns in the data, resulting in higher prediction errors and negative R-squared values that indicate their inability to outperform a simple mean-based model. The strong performance of the MLP Regressor could be attributed to its ability to capture non-linear relationships in the dataset, making it a more suitable choice for the given task. It can generalize to many shapes like sinusoidal, parabolic, and exponential, for example, allowing detection of more intricate patterns.

The MLP Regressor, deemed the most accurate model, is also practical. After further testing by passing in new data, the model provided close to accurate ESG scores for most cases, but was sometimes off by a lot. Overall, the model delivered believable results, not creating negative numbers or numbers not associated with ESG.

Conclusion

This study has demonstrated the potential of machine learning techniques in predicting Environmental, Social, and Governance (ESG) scores based on comprehensive financial data. By leveraging datasets from Kaggle and Yahoo Finance and using the feature engineering process, we identified key financial metrics that exhibit a strong relationship with ESG performance.[7,8]

The Multilayer Perceptron Regressor emerged as the most accurate model, outperforming traditional linear regression models and variants such as Lasso, Ridge, and Elastic Net regression. Its ability to capture complex, non-linear patterns within the data enabled it to achieve a low mean squared error of 0.145 and a high R-squared value of 0.989, indicating predictive power and variance explanation.

While the model exhibited promising results, there is still room for improvement. The small dataset does not allow for generalization to all industries in the world, but only to the information technology sector. Future research could incorporate larger datasets, encompassing a broader range of companies and industries, to enhance the model's generalizability. For example, including more industries like companies in the Dow Jones Industrial Average will generalize the findings even further. Additionally, exploring ensemble techniques or incorporating qualitative factors may further refine the model's predictive capabilities. Furthermore, using more advanced neural networks like Deep Neural Networks might result in stronger results because they can detect more nuanced patterns in the data.

This study's findings have significant implications for investors, analysts, and companies. By providing a quantitative framework for assessing ESG performance, this research enables more informed decision-making and facilitates the alignment of financial strategies with environmental, social, and governance considerations. Ultimately, this work contrib-

DOI: 10.36838/v7i10.12

utes to the broader goal of promoting sustainable and ethical business practices, underscoring the potential of data-driven approaches to driving positive change.

Acknowledgments

I want to thank my mentor, John Lee, for guiding me to complete this research. He helped me learn more about finance, emphasizing each financial metric's importance. He steered me away from rabbit holes and kept me on the right track. He provided me with motivation and constant support throughout this project. I am truly grateful for the assistance and encouragement throughout this study.

References

1. Op 5 ESG Investment Trends in 2023. KnowESG. https://www.knowesg.com/featured-article/top-5-esg-investment-trends-in-2023 (accessed 2024-12-09).
2. Adame, L.; Soto, R.; Ballinez, R.; Prieto, M.; Villaseñor, M.; Boni, F. ESG Evaluation Criteria: Environmental, Social and Governance Analysis; HR Ratings: December 2021. Available at: https://www.hrratings.com/docs/metodologia/ESG_Evaluation_Criteria_December_2021.pdf (accessed Dec 2024).
3. Kim, M.-S.; Lee, H.-J.; Jung, S.-H.; Lee, S.-Y.; Kim, H.-J. Predicting Future ESG Performance Using Past Corporate Financial Information: Application of Deep Neural Networks. J. Intell. Inf. Syst. 2023, 29 (2), 85–100. https://doi.org/10.13088/JIIS.2023.29.2.085. (1)
4. Lin, H.-Y.; Hsu, B.-W. Empirical Study of ESG Score Prediction Through Machine Learning—A Case of Non-Financial Companies in Taiwan. Sustainability 2023, 15, 14106. https://doi.org/10.3390/su151914106. (2)
5. Adeneye, Y.; Kammoun, I.; Aqilah, S. Capital Structure and Speed of Adjustment: The Impact of Environmental, Social, and Governance (ESG) Performance. Sustainability Accounting, Management and Policy Journal 2022, 14. https://doi.org/10.1108/SAMPJ-01-2022-0060. (3)
6. Moody's Analytics. ESG Score Predictor. https://www.economy.com/products/esg/esg-score-predictor (accessed Apr 6, 2024). (4)
7. Kaggle. S&P 500 ESG Risk Ratings. https://www.kaggle.com/datasets/pritish509/s-and-p-500-esg-risk-ratings (accessed Apr 6, 2024). (5)
8. Arrousi, R. Yfinance. https://pypi.org/project/yfinance/ (accessed Apr 6, 2024). (6)
9. D'Amato, V.; D'Ecclesia, R.; Levantesi, S. Fundamental Ratios as Predictors of ESG Scores: A Machine Learning Approach. Decisions Econ. Finan. 2021, 44, 1087–1110. https://doi.org/10.1007/s10203-021-00364-5.
10. Huber, B. M.; Comstock, M. ESG Reports and Ratings: What They Are, Why They Matter. The Harvard Law School Forum on Corporate Governance. https://corpgov.law.harvard.edu/2017/07/27/esg-reports-and-ratings-what-they-are-why-they-matter/ (accessed Apr 6, 2024). (7)
11. A.M.O. Hmouda; G. Orzes; Sauer, P. C.; Molinaro, M. Determinants of Environmental, Social and Governance Scores: Evidence from the Electric Power Supply Chains. Journal of Cleaner Production 2024, 471, 143372–143372. https://doi.org/10.1016/j.jclepro.2024.143372.
12. D'Amato, V.; D'Ecclesia, R.; Levantesi, S. ESG Score Prediction Through Random Forest Algorithm. Comput. Manag. Sci. 2022, 19, 347–373. https://doi.org/10.1007/s10287-021-00419-3.
13. Current Ratio: Assess Your Ability to Pay Short-Term Liabilities. Allianz Trade Corporate. https://www.allianz-trade.com/en_US/insights/current-ratio.html.
14. Fernando, J. Earnings Per Share (EPS): What It Means and How to Calculate It. Investopedia. https://www.investopedia.com/terms/e/eps.asp.
15. Fernando, J. Free Cash Flow (FCF). Investopedia. https://www.investopedia.com/terms/f/freecashflow.asp.
16. Tuovila, A. Operating Cash Flow (OCF) Definition. Investopedia. https://www.investopedia.com/terms/o/operatingcashflow.as
17. What is Multilayer Perceptron? Deepchecks. https://www.deepchecks.com/glossary/multilayer-perceptron/.
18. Introduction to Simple Linear Regression 1. Regression Equation. https://murraylax.org/eco307/notes/regression_intro.pdf.
19. Agrawal, S. Feature Selection Using Lasso Regression. Medium. https://medium.com/@agrawalsam1997/feature-selection-using-lasso-regression-10f49c973f08.
20. IBM. What is regression? | IBM. www.ibm.com. https://www.ibm.com/topics/lasso-regression.
21. Murel, J. What is ridge regression? | IBM. www.ibm.com. https://www.ibm.com/topics/ridge-regression.
22. Brownlee, J. How to Develop Elastic Net Regression Models in Python. Machine Learning Mastery. https://machinelearningmastery.com/elastic-net-regression-in-python/.
23. Kumar, V. Regularization Methods: Lasso (L1), Ridge (L2) and Elastic Net. Medium. https://medium.com/@ipvikas/regularization-methods-lasso-l1-ridge-l2-and-elastic-net-2bf5bb35d5ee (accessed 2024-12-09)
24. https://www.graphpad.com/support/faq/how-can-rsup-2sup-be-negative/#:~:text=R2%20will%20be%20negative%20when%20the%20line%20or%20curve,different%20data%20set)%2C%20or%20
25. De Lucia, C.; Pazienza, P.; Bartlett, M. Does Good ESG Lead to Better Financial Performances by Firms? Machine Learning and Logistic Regression Models of Public Enterprises in Europe. Sustainability 2020, 12, 5317. https://doi.org/10.3390/su12135317.
26. https://github.com/aravindkrishna2008/ESG-Research-Paper/blob/main/Training_%2B_Graphs.ipynb

Author

Aravindkrishna Arivudaianmbi is a sophomore at Lynbrook High School in San Jose, California. He is interested in machine learning and artificial intelligence and wants to use his knowledge to generate new and innovative ideas for using machine learning to predict the future better.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Securing Generative AI: Homomorphic Encryption, Differential Privacy, and Federated Learning in Key Industries

Avani Thakur

Saint Francis High School, 1885 Miramonte Avenue, Mountain View, CA, 94040, US.A.; avani.thakur22@gmail.com

ABSTRACT: Generative AI (GAI) has revolutionized industries such as marketing, healthcare, and finance by enabling advanced data analysis through Large-Language Models (LLMs) and Transformers. However, their reliance on vast amounts of training data raises significant privacy concerns—models can unintentionally memorize and regurgitate user-specific text, leak proprietary health records via extraction attacks, and be manipulated through prompt-injection exploits, transforming privacy from an option to a necessity. This paper investigates three key Privacy-Preserving Mechanisms (PPMs) that can be applied to Machine Learning (ML) models to prevent sensitive data memorization while minimizing impacts on performance. This analytical review, therefore, examines which Privacy-Preserving Mechanisms (PPMs) can be applied to Machine Learning (ML) models to prevent the memorization of sensitive training data while minimizing adverse effects on model performance. It discusses how these methods protect sensitive information while balancing the trade-offs between privacy guarantees and model utility, and examines practical implementations across real-world applications. The review also identifies current challenges and potential future research directions, providing valuable insights for academic researchers and industry practitioners working to secure data-driven technologies. This is particularly important in today's world because AI is being implemented in sensitive domains with innumerable ethical implications. So, understanding why these challenges have arisen and their potential solutions is crucial before the applications of AI continue to increase and advance.

KEYWORDS: Keywords: Machine Learning, Artificial Intelligence, Cybersecurity, Homomorphic Encryption, Differential Privacy, Federated Learning, Large Language Models, Transformers.

Introduction

In 2024, X (previously Twitter) experienced a massive data breach, compromising nearly 200 million user records.[1] In February of 2024, the biggest healthcare breach occurred in Change Healthcare, Inc., where the ransomware group, BlackCat/ALPHV, extracted the protected health information of over 100 million individuals.[2] Between February and May of 2024, Evolve Bank & Trust was also a victim of a ransomware attack that compromised the personal data of around 8 million customers.[3] These high-profile incidents reveal the escalating urgency of data privacy, as breaches are no longer isolated failures but systemic breakdowns–eroding customer trust, exposing companies to severe regulatory and financial liability, and triggering multi-million-dollar fines under regulations like the General Data Protection Regulation (GDPR). The GDPR is a European law that grants individuals control over their personal data and imposes penalties of up to 4 percent of global annual turnover for non-compliance.[4] Moreover, the unprecedented pace of AI integration—where millions of user interactions are processed in real time by LLMs—dramatically amplifies the attack surface, enabling criminals to exploit unprotected data at scale and compelling organizations to implement robust privacy safeguards immediately to prevent the next breach from causing irreversible harm. This growing urgency is precisely what motivates us to conduct this systematic survey of Homomorphic Encryption, Differential Privacy, and Federated Learning—to map the current landscape of defenses, identify critical gaps, and chart a path toward comprehensive, end-to-end privacy for AI-driven systems.

ML models heavily rely on large amounts of training data to perform complex tasks.[5] However, this dependency raises a significant concern: the potential memorization of private or sensitive data during training.[6] When exposed to extensive training datasets, such models may unintentionally retain specific details from the data, such as personal information, proprietary content, or other confidential material.[7] Prior research[8] has even found a linear relationship between the frequency of specific sequences in a training dataset and the likelihood of those sequences being memorized by the model, raising serious privacy concerns.

Privacy and security, while related, address different aspects of data protection. Often, these two words are used interchangeably, yet they refer to different aspects of data protection. Privacy pertains to the rights and controls that an individual has over their personal information – how it is collected, used, and then shared. Security, on the other hand, refers to the measures and protocols used to protect this data from unauthorized access, breaches, and threats.[9] So, in simplistic terms, privacy focuses on governing the use of personal data, while security concentrates on safeguarding against potential dangers. Yet, privacy remains a complex subject, often ambiguous in concept, and is frequently described as "a concept in disarray."[10]

DOI: 10.36838/v7i10.18

Despite numerous point solutions—such as deploying homomorphic encryption to protect data in transit, injecting noise during training under differential privacy, or aggregating local updates via federated learning—real-world deployments of LLMs continue to suffer from fragmented defenses. As the Wall Street Journal has reported, companies are now grappling with conversational attacks, prompt injections, and the inadvertent leakage of proprietary code or corporate secrets into publicly accessible LLMs. Moreover, the rapid commoditization of these models means adversaries can cheaply spin up clones with fewer governance guardrails, further widening the attack surface.[11]

What is missing is a unified framework that secures data end-to-end: remaining encrypted while in motion, never revealing sensitive details during model training, and still delivering high-quality generative performance at inference. To date, no single approach has simultaneously addressed these three domains—transit, training, and inference—without imposing untenable computational costs or degrading model utility. We therefore need a coherent synthesis of HE, DP, and FL tailored to LLMs and Transformers.

This paper presents a systematic review of various Privacy-Preserving Mechanisms that can be applied to ML models, with a focus on Generative AI tools in high-stakes industries, to prevent the memorization of training data while minimizing the impact on model performance. We structure our survey along two dimensions—first by PPM techniques (HE, DP, FL) and then by application domains (marketing, healthcare, and finance), noting that many studies span both categories and may be discussed under multiple headings. This paper will compare various methodologies and assess their efficiency as well as their strengths and weaknesses.[12] At the end, we also propose a new system design that combines all three PPMs discussed in this review, for more robust privacy. This survey is especially relevant in today's world, where there is a growing reliance on AI tools, as privacy risks within such ML models do not solely impact individuals but organizations as a whole. This risk is especially pronounced in sectors like marketing, healthcare, and finance, as it directly impacts customer trust, which, if broken, could dismantle entire companies and damage reputations.[13]

Related Work

Data is continuously being collected from users through interconnected devices and sensors. Through such data collection, users risk exposing sensitive data to untrusted individuals and/or companies. Such entities can then analyze the data to extract private information, which is then used maliciously (i.e., selling information to third parties). Due to an increasing number of hacks and breaches, privacy is a growing concern for users, and many want to know how to protect their data. To address this, Privacy-Preserving Mechanisms (PPMs) have been introduced. PPMs are mainly used to preserve the privacy of a user, but there can be some complications, as Cunha, Mendes, and Vilela have discussed in their paper.[14]

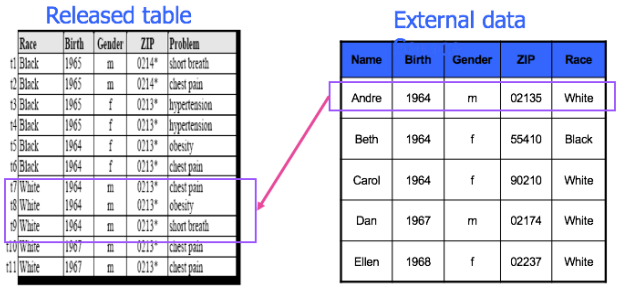
Cryptography:

Cryptography is a method used to conceal information so that unintended parties cannot access it. There are two main types: symmetric and asymmetric cryptography. In symmetric cryptography, both the sender and the receiver share one key that is used for encrypting and decrypting a message. In contrast, asymmetric cryptography uses two different keys–a public key and a private key. The public key can be freely shared and used by anyone to encrypt a message, but only the holder of the private key can decrypt it. This means that only the holder of the private key can access the original information, while everyone else can still encrypt. This provides an extra layer of privacy, keeping the information safe from unauthorized access.[15]

K-Anonymization:

K-anonymization is a technique that protects individual privacy by ensuring that every record in a dataset is indistinguishable from at least k-1 other records based on key identifying attributes. For instance, if k is set to 5, then any record's combination of identifying details will appear in at least five records, making it difficult to pinpoint a single individual. This is achieved by generalizing or suppressing specific details, which hides a person's unique identity within a group while still preserving the overall usefulness of the data for analysis.[16]

Table 1: This table demonstrates k-Anonymity by ensuring each record shares its quasi-identifiers (e.g., birth date, gender, ZIP code) with at least k – 1 others, forming equivalence classes that prevent singling out individuals. By generalizing or suppressing these identifiers, linking to an external dataset no longer reveals sensitive information about any single person.[17]



This table on the left is a generalized data set with identifiers. Quasi identifiers are attributes such as Date of Birth, Gender, and Zip Code, which can identify 87% of the U.S. population. These can be used to link anonymized datasets with real datasets, which brings us back to Table 1. The table on the right is an unrelated dataset, for example, a voter database. By linking the generalized dataset with the real dataset, it makes it impossible to figure out Andre's problem through using K-Anonymization (Table 1).

Privacy-Preserving Mechanisms:

Privacy-Preserving Mechanisms modify or decentralize data handling to prevent leakage of PII while retaining analytical value. This review focuses on three complementary approaches—Homomorphic Encryption, which keeps data encrypted throughout computation; Differential Privacy, which adds calibrated noise to outputs to bound individual disclosure; and Federated Learning, which trains shared models on decentralized data without moving raw records. Each method embodies a distinct privacy–utility trade-off—HE maximizes confidentiality at the cost of computational overhead, DP provides formal leakage guarantees with potentially reduced accuracy,

and FL balances collaboration with minimal data exposure via parameter sharing. These trade-offs and their implications for real-world deployments in marketing, healthcare, and finance are explored in depth later in the paper.

Homomorphic Encryption

Homomorphic Encryption (HE) is a powerful PPM that, when fully and correctly implemented, has the potential to revolutionize cybersecurity.[18] This PPM allows various operations to be performed directly on encrypted data without ever decrypting it.[19] When the data is eventually decrypted, the same operations will have been applied to the decrypted data, all without the intermediary ever accessing or seeing the original unencrypted information. This ensures that sensitive data, such as Personally Identifiable Information (PII), remains secure even when shared with external entities for analysis. By preventing intermediaries from accessing the raw data, HE minimizes privacy risks and protects user confidentiality.

Homomorphic encryption's promise is tempered by its complexity and the different types of HE available: Partial Homomorphic Encryption (PHE), Somewhat Homomorphic Encryption (SWHE), and Full Homomorphic Encryption (FHE). The reason lies in its complexity and the types of HE available: Partial Homomorphic Encryption (PHE), Somewhat Homomorphic Encryption (SWHE), and Full Homomorphic Encryption (FHE). Among these, FHE stands out as the most secure because it enables various operations to be performed entirely on encrypted data without ever decrypting it into plaintext. This makes FHE ideal for securely handling sensitive information, even with untrusted individuals. However, practical FHE implementations remain limited.[20]

In 2009, Craig Gentry, a PhD graduate from Stanford University, developed the first successful FHE scheme based on mathematical lattices.[21] Despite its groundbreaking potential, FHE faces challenges in real-world applications due to inefficiencies. Specifically, as more operations are performed on encrypted data, FHE accumulates noise, making the final decryption process inaccurate or even impossible. To address this, Gentry introduced the concept of a secret key to reduce noise. Instead of using the secret key directly, he proposed encrypting the key itself, allowing it to mitigate noise during decryption. This innovation significantly improved the accuracy and reliability of FHE, making it more viable, though still not as efficient as PHE or SWHE, as FHE operations are considerably slower, often tens to thousands of times slower than equivalent plaintext computations. Even though FHE hardware accelerators are being developed to address these speed limitations,[22] their adoption remains limited, making software-based optimizations the more practical approach for now.

Homomorphic Encryption in LLMs:

With advanced natural language processing (NLP) capabilities, LLMs can generate human-like text, handle customer inquiries, create personalized content, and assist in decision-making—all at a fraction of the time and expense required for human labor.[23]

However, these AI technologies come with significant drawbacks that must be responsibly addressed. This is where privacy-preserving mechanisms can be employed, with homomorphic encryption being one of the most promising solutions. Applying HE to LLMs poses three main challenges. First, modern LLMs contain billions of parameters, so encrypting every weight and activation vastly increases both memory usage and computational cost. Second, core operations such as Softmax (involving exponentials and divisions) and LayerNorm (involving means and square roots) are non-polynomial and thus incompatible with HE's addition–multiplication framework. Third, managing the accumulated noise in ciphertexts to remain below decryption thresholds often requires costly bootstrapping, further degrading performance.[24] However, to successfully implement a Fully Homomorphic Encryption (FHE) scheme on an LLM, the model must be reconfigured to support secure computations and integrate PPMs, enabling operations to be performed on encrypted data without compromising its security.[25]

This can be achieved by transforming a typical AI or Large Language Model (LLM) into a privacy-preserving computational model. In data-driven and model-driven computation scenarios, two primary components exist: the user's data, such as a query or prompt, and the trained LLM model. Additionally, multiple parties are typically involved, including the user (client), the computational model, and the model owner. The proposed transformation ensures privacy by encoding and embedding user data into polynomial spaces while representing the LLM model as polynomial functions. To encode the user's data, feature extraction and vectorization techniques are employed using tools like Word2Vec or GloVe to create secure representations of the input data. Meanwhile, the model itself undergoes compression to reduce its size. Compressed models are then converted into polynomial form to facilitate secure computations on encrypted data.

The computational workflow follows a systematic process. First, features from the user's input are first extracted and then encoded into polynomial spaces. Cryptographic keys are then generated to secure operations, followed by encrypting the encoded data. The compressed and arithmetic model processes this encrypted data securely through privacy-preserving delegation, enabling computations without exposing sensitive information. The final output is then decrypted to present results to the user while maintaining privacy throughout. Neural networks naturally rely on complex polynomial computations, aligning seamlessly with this approach.

It supports a robust system where user data is secure while computations remain efficient. This methodology provides a versatile framework for converting LLMs into privacy-preserving models tailored to specific use cases and computational requirements. [26]

Homomorphic Encryption in Transformers:

Transformers, with their advanced capabilities in word prediction, sentence completion, etc., have gained popularity in the last few years. However, their reliance on large amounts of training data raises significant privacy concerns, just like LLMs, due to memorization risks and the inadvertent release

DOI: 10.36838/v7i10.18

of sensitive information. This is why, in recent years, HE has been applied to transformers in an attempt to increase privacy without degrading the functionality of these models. In this subsection, we examine two approaches that both apply FHE schemes to transformers, focusing on the tradeoff between privacy and utility.[27]

The first method starts by finding the parts of a transformer model that are hard to secure with HE, which usually involves functions with square roots. These functions are complicated and not easy to convert into simple polynomial forms, which are necessary for secure computations. To overcome this, researchers redesigned the model by replacing these complex functions with simpler, HE-friendly alternatives. For example, they introduced a new attention mechanism that uses an activation function (σ) along with a length scaling factor to stabilize the model. They also added extra training steps and loss functions to narrow the range of values in the model, making it easier to replace non-polynomial functions with polynomial approximations. This three-step process—modifying the architecture, retraining with range minimization, and finally substituting with polynomial approximations—creates a secure, HE-friendly transformer model. The results showed that the modified, secure transformer worked almost as well as the original model. In natural language tasks, the new attention mechanism prevented the model's weights from getting too big and helped it learn smoothly. In image classification tasks, such as on the CIFAR-10 and CIFAR-100 datasets, the new method improved accuracy significantly, by over 11% on CIFAR-10 and about 2.5% on CIFAR-100. Additionally, when the model was used for secure inference with HE, it performed efficiently, even setting a competitive benchmark compared to similar models. Overall, the results indicate that it is possible to build a secure transformer model that maintains high performance while protecting sensitive data.[28]

In the second approach, a small version of BERT (called BERTtiny) was run using FHE so that sensitive data remains encrypted even while the model processes it. In this setup, a client tokenizes a sentence and creates token embeddings, which are then encrypted and sent to an "honest-but-curious" server. This server uses FHE based on the RNS-CKKS scheme (an FHE scheme that calculates approximations on real or complex numbers) to run the transformer encoder layers, a pooling layer, and a classification layer—all without ever decrypting the data, thus preserving privacy. The results of this approach, however, were not as successful as the first approach, suggesting that future improvements could include scaling their architecture to support larger BERT models, using more accurate polynomial approximations of activation functions, or designing neural networks that produce well-separated outputs. This will allow less precise HE computations to be used, leading to faster and more efficient processing. [29]

Homomorphic Encryption in Marketing:

Customer trust is a cornerstone of successful businesses and products, as it enables organizations to build a loyal customer base and drive sales.[30] This is particularly critical in the marketing sector, where trust directly influences the reputation of an organization. If trust is broken, it can jeopardize an entire company's future. However, the introduction of LLMs and Transformers has raised new privacy concerns regarding how and where customer data is handled. This transparency is essential for customers, as they may choose to avoid certain products or services if they feel their data is not being properly protected.[31]

The secure SMS-targeting framework is designed to send messages to a specific audience. It has four main components: the owner of consumption data, the user of consumption data, a third-party computing center, and an advertising operator. The e-commerce entity starts by using encryption techniques to encrypt the consumption data. Then, this encrypted data is stored in a protected location from which the third-party computing center retrieves the ciphertext for further processing. Then, when the merchant sends the personalized SMS marketing messages, they just input their desired customer categories and marketing content. Through this framework, merchants will be unable to access individual information, allowing for personalized messages, without putting the privacy of users in jeopardy (Figure 1).

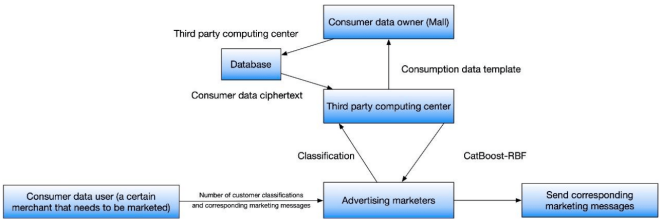


Figure 1: Overall, this approach provides a robust framework that allows marketers to successfully send personalized SMS messages while still maintaining customer privacy.

Homomorphic Encryption in Healthcare:

The healthcare sector processes vast amounts of patient data, including highly sensitive PII such as full names, health problems, addresses, etc. When this information is shared between facilities, significant privacy concerns arise regarding who has access to what data. Regulations like the Health Insurance Portability and Accountability Act (HIPAA) require that patient data remain unreadable by unauthorized third parties.[33] HE and FHE offer a promising solution by enabling third parties to perform computations on encrypted data—allowing analysis to be done without ever accessing the raw data—thereby ensuring compliance with privacy regulations.[34] Although HE faces practical challenges such as excess noise, it has already been studied in various healthcare scenarios.

In this approach, researchers said that specialized libraries and toolkits—like Microsoft's SEAL and IBM's HELib—make it much easier for developers to integrate HE into their applications by providing prebuilt functions and APIs that optimize various encryption schemes. Such resources simplify complex tasks such as performing secure computations without decryption. Furthermore, international programs in this industry showcase the real-world benefits of HE. For example, encryption systems such as CryptDB and Mylar are used to secure sensitive patient data, ensuring that even when the data is stored or transmitted, it remains unreadable to unauthorized users. This means that healthcare providers and researchers

can analyze confidential data—such as patient records and diagnostic results—without ever exposing the raw information, helping to meet strict regulations like HIPAA and General Data Protection Regulation (GDPR).[35]

Homomorphic Encryption in Finance:

In the finance sector, huge amounts of data are tracked and analyzed that contain PII, such as bank account numbers, SSNs, etc. However, with the rise of LLMs and Transformers in this sector, privacy risks have started to emerge.[36] At the same time, some approaches have already been proposed on how to make such models secure using HE in this industry. In this subsection, we will examine an approach that has been proposed.

FinancialCloud leverages HE to allow pricing services to perform computations directly on encrypted data, ensuring that sensitive financial information is never exposed during processing. In this framework, an e-commerce entity encrypts its consumption data, which is then stored securely. When a merchant needs to send targeted marketing messages, the pricing service retrieves the encrypted data and performs all necessary computations without decrypting it. European option pricing is a key example highlighted in the study, where a binomial tree method is used. Here, the system first computes all possible future stock prices in an encrypted domain, applies HE-based arithmetic (addition, multiplication, division) to evaluate option values at terminal nodes, and then works backward to determine the option price. Experimental results demonstrate that while HE significantly increases computation time compared to non-encrypted methods, it also provides strong data security. The improved algorithm achieves better performance, reducing the computation time from 1835.17 seconds to 1375.59 seconds.[37]

■ Differential Privacy

Differential privacy (DP) was first formalized in 2006 by Cynthia Dwork and her colleagues, who introduced rigorous mathematical frameworks to measure and guarantee privacy protection.[38] At its core, DP balances the trade-off between privacy and utility by carefully calibrating noise to provide quantifiable privacy guarantees while preserving as much data accuracy as possible. Over the years, DP has been adopted by major organizations, including Apple and Google, to safeguard user data while still enabling valuable insights. DP has the potential to transform privacy across various domains. This PPM ensures that sensitive information within a dataset remains protected, even when the dataset is queried or analyzed.[39] By introducing a controlled amount of random noise to the result of queries, DP obscures the contribution of a single data point, making it close to impossible to identify individual records. This enables organizations to extract useful insights from large datasets—such as those containing PII, healthcare records, or financial data—without exposing sensitive details about individuals.[40]

Differential privacy is not the default solution for all privacy concerns because of its inherent trade-offs between privacy and utility. The answer lies in the inherent trade-offs between privacy and utility. While DP can effectively shield individual data points, the added noise can sometimes reduce the accuracy of the results. The more privacy that is preserved, the more noise is introduced, which can then potentially compromise the dataset's utility for certain applications. Striking the right balance between privacy and data usability requires careful calibration, often tailored to the specific context and sensitivity of the data.

Despite its promises,challenges remain in scaling DP to complex use cases, such as real-time data streams or applications requiring high accuracy. Moreover, its implementation can be computationally expensive and requires specialized expertise to integrate into large-scale analytics systems.

This is why Federated Learning is often paired with Differential Privacy solutions to create more secure models. FL is another privacy-preserving mechanism that will be discussed in greater detail later; however, understanding FL is crucial for comprehending some of the DP approaches mentioned in this section. In FL, multiple devices collaboratively train a model without ever sharing their raw data. Instead, each device trains a local model on its private data and then sends its encrypted model parameters to a central server. The server then decrypts, averages, and integrates those parameters to update a centralized model. This updated model's parameters are then sent back to the local devices, which further train it on their private data. This iterative process continues until the centralized model is fully trained, ensuring that sensitive data remains on local devices throughout the process.[41]

Differential Privacy in LLMs:

To provide actionable guidance for deploying differential privacy in real-world LLM fine-tuning—where practitioners must balance privacy budgets against model performance under diverse operational constraints—we compare four DP approaches that have been applied to improving privacy in LLMs: BFL-LLM,[42] EW-Tune,[43] Whispered Tuning,[44] and variants of DP-SGD (ELS and ULS).[45] Throughout, we analyze each method's position on the privacy–utility spectrum to identify the most practical balance for different deployment scenarios.

The BFL-LLM model integrates differential privacy, federated learning, blockchain technology, and secure multi-party computation (SMPC). It starts by encoding raw user data at the device layer to secure cross-silo FL environments, ensuring that the raw data is protected before it gets transmitted for local model training. Then it employs secret sharing and SMPC at the edge layer, where local models are trained to prevent any single entity from accessing complete user data. Then, in the cloud layer, blockchain technology is employed to encrypt and aggregate the locally trained models. Results showed that this model balanced privacy with data utility through multi-feature encoding and adjustable noise parameters. This was demonstrated on various datasets as it rejected poisoned submissions, increasing the global model accuracy.

EW-Tune, on the other hand, offers DP in situations with a finite number of compositions–i.e., iterations of Stochastic Gradient Descent (SGD) (ML algorithm that optimizes models by using small batches of data to update parameters)–rather than asymptotic privacy bounds. This requires less noise during SGD, which then directly improves model accuracy.

DOI: 10.36838/v7i10.18

This model offers a clearer improvement by reducing the noise introduced during SGD by up to 5.6% and also gaining model accuracy by 1.1%.

The Whispered Tuning model combines PII redaction, differential privacy, output filtering, and architectural improvements. In order for their approach to work, they had to first manufacture their own dataset with fake PII. They created two hundred and two stories with PII (full name, SSN, credit card number, phone number, etc.). They begin by first fine-tuning DistilBERT to redact PII from their dataset in order to create a new redacted dataset with only the PII. Then they fine-tuned two pre-trained models on this redacted dataset to adapt them to domain-specific language nuances. They continued by further fine-tuning using the Supervised Fine-Tuning Trainer (SFTTrainer). Next, masked models were fine-tuned on the unredacted dataset using DP techniques through the Opacus library. This process utilizes an "Epsilon Dial" (ε) to customize privacy budgets, where a lower ε level offers stronger privacy with reduced data utility, but a higher ε level offers more utility for roles that do not require as much privacy. The final phase is the "self-reflection filter" that scans and replaces any leaked PII using artificial and non-sensitive data from the Faker library. In this framework, the ClearView model provides highly engaging and personalized responses while leaking PII, whereas the SecureNLP models generate more generic outputs that better protect user privacy, which is not a great tradeoff between privacy and utility.

Finally, the DP-SGD approach is evaluated in two variants–Example-Level Sampling (ELS) and User-Level Sampling (ULS). DP-SGD-ELS protects individual examples using a key group size parameter (GELS) that determines the number of examples that should be used per computation. If GELS is set too low, the resulting dataset may also be too small and have less diversity because many users contribute more examples than the allowed group size. But, if GELS is set too high, more noise needs to be added to maintain privacy for users who contribute many examples, which can then degrade the accuracy of the model. So, the optimal setting for GELS should be close to the median user dataset size. On the other hand, DP-SGD-ULS protects all contributions of a user by tuning group size (GULS) and cohort size (M). The optimal setting for these two key parameters is to increase M up to some limit first and then increase GULS, with GULS staying much lower than M. These approaches on DP-SGD, showed that by adjusting group and cohort sizes based on budget and privacy parameters (ε), minimizing utility degradation while still maintaining privacy is attainable.

Differential Privacy in Transformers:

In this subsection, we compare three DP approaches that have been applied to Transformers–Phantom Clipping, Re-Attention Mechanism,[46] and Privately Pre-Training Transformers[47]–focusing once again, on the trade-offs between privacy and utility.

To start, Phantom Clipping calculates the per-sample gradient norm needed for DP-SGD in a way that mimics the standard DP-SGD process, so it naturally retains the same privacy guarantees. It determines how much each data sample influences the model's update without compromising privacy, and it does so using a predictable amount of memory for the embedding layer. This method avoids creating separate gradients for each sample (which is what older methods like Ghost Clipping do), thus speeding up the process and reducing memory usage. The results show that the trade-off is that careful tuning is needed to ensure that the reduced noise still offers strong privacy protection without compromising utility. The benefits of Phantom Clipping are most noticeable in smaller, resource-constrained models, as this approach provides efficiency, which is especially beneficial for such models, where keeping data local is crucial for privacy. However, the approximations it uses may introduce small inaccuracies that can slightly affect overall model performance. For larger models, the efficiency gains might be less significant, and fine-tuning becomes more complex, potentially affecting the delicate balance between privacy and performance.

Next, the Re-Attention Mechanism changes how their model processes input compared to standard DP-SGD, but these changes do not cause samples in a mini-batch to depend on each other, which is important because each sample's contribution must remain isolated, so any noise added to protect one sample does not inadvertently impact the others. This independence helps ensure that the privacy guarantees apply uniformly to each data point in the mini-batch. This means that the sensitivity of the model–how much one sample influences the output–can still be controlled using gradient clipping as in DP-SGD. Additionally, this method uses techniques from Bayesian deep learning (a framework that incorporates uncertainty when training models) to approximate a difficult-to-compute variance term in a "black-box" way, meaning these techniques can be swapped with others if needed, as long as they effectively estimate the variance. This model's results help strike a better balance between protecting data and maintaining model accuracy as the model reduces distraction with the additional noise that comes along with stronger privacy, which is what can hurt the performance of the model. This model stays stable and learns consistently even when extra noise is added for privacy, which is crucial since higher privacy often means more noise and potential learning issues.

Finally, the last approach focuses on privately pre-training a model rather than fine-tuning a publicly pre-trained model to see if that yields a better tradeoff between privacy and utility. The results of this model found that using a specialized DP technique, DP-SP, as a regularizer during pre-training significantly improves performance (about 13% better in optimal settings). Additionally, models fine-tuned with DP techniques show that even with a drop in pre-training accuracy, the final fine-tuned performance on tasks remains close to that of non-private models. However, experiments revealed that when randomly selected token spans are replaced with a special mask token, it is highly effective at preventing the model from memorizing and outputting training data verbatim, while models using prefix training (a fine-tuning technique where a small, trainable "prefix" of continuous vectors is added to the input) leak more training data. When their approach is fully

implemented, memorization is reduced considerably, especially when combined with deduplication techniques.

All three studies corroborate that stronger privacy invariably introduces some performance degradation, but they differ in their strategy: Phantom Clipping prioritizes efficiency in small-scale setups, Re-Attention Mechanism balances noise injection with Bayesian approximations for stability, and Privately Pre-Training emphasizes end-to-end privacy at the cost of significant compute overhead. Yet, each approach's empirical validations span different model sizes and tasks, underscoring the need for future work to benchmark these methods under uniform criteria to guide real-world adoption.

Differential Privacy in Marketing:

In this subsection, a DP application in a bid optimization is examined to shed light on the broader digital marketing trends.

This approach analyzes three different techniques of DP. Noise addition is a method that adds random noise to each bid using a function that generates values based on a specified mean and variance. By modifying the original bid amounts with this random noise, the exact bid values are obscured. This ensures that while the bidding strategy remains competitive, the precise information about each bid is hidden, protecting sensitive user data. Similarly, in data perturbation, the original bid data is transformed using a perturbation function. This function introduces uncertainty into the data representation by altering the bids in a controlled manner. The result is a set of modified bids that maintain the overall statistical properties needed for effective bidding but make it more difficult for adversaries to extract specific, sensitive details from the data. Finally, anonymization is a technique that further protects privacy by aggregating bids across similar user profiles. Instead of dealing with individual bid values, the data is grouped into clusters where each group contains similar bids. This process effectively hides the identity of individual bids by making it impossible to trace any specific bid back to a particular user, thereby enhancing overall data privacy. In this approach, reducing noise-induced performance loss while still ensuring strong privacy guarantees is the optimal setting. Overall, despite the inherent trade-off where stronger privacy measures typically reduce accuracy compared to non-private baselines, the evaluated approaches effectively mitigate this loss, achieving competitive bid optimization performance while preserving user privacy.[48]

Differential Privacy in Healthcare:

The application of DP in this sector has become increasingly important to safeguard sensitive information while preserving the overall usefulness of the data. In this subsection, two different DP approaches will be examined to analyze the tradeoff between privacy and utility.

The first approach focuses on cohort identification for clinical trials, where the goal is to count eligible patients without exposing sensitive patient data. Traditional methods only offer approximate privacy guarantees. On the other hand, this new method modifies the exponential mechanism to incorporate user preferences, favoring more accurate outputs while still protecting privacy. It also introduces predefined query budgets and sets the overall privacy budget (ε) to match existing standards, aiming to balance accuracy with strong privacy. The results used predefined query budgets to manage noise based on expected counts, and it proposes a way to set the overall privacy budget (ε) to match the protection level of I2B2, even though it does not fully address all challenges of differential privacy.[49]

The second approach applies DP in an FL context to predict adverse drug reactions (ADR) using the LCED dataset and in-hospital mortality using the MIMIC III dataset. Here, three classifiers—perceptron, support vector machine (SVM), and logistic regression—are trained using 5-fold cross-validation across 10 sites. Their experiments measured the trade-off between privacy and utility by varying ε. The results for this approach show that as ε increases (privacy relaxes), the model's F1 score improves. Even though FL performance is comparable to centralized learning, applying DP in real-world healthcare settings with a limited number of sites leads to a significant loss in utility. Overall, both approaches illustrate the current challenge of balancing robust privacy protection with high model performance in sensitive healthcare applications (Figure 2).[50]

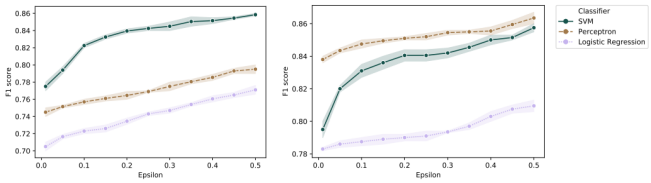


Figure 2: F1 scores measure utility for adverse drug reaction and in-hospital mortality prediction under centralized and federated learning. Across both LCED and MIMIC datasets—and for all three classifiers—applying ε-differential privacy yields a monotonic increase in F1 as ε rises (i.e., as privacy constraints relax), with FL models closely matching centralized baselines at higher ε values.

Differential Privacy in Finance:

In this subsection, three different approaches will be analyzed to understand how the tradeoff between privacy and utility can best be balanced.

The first approach is the DP-LSTM model, which is designed for predicting S&P 500 stock prices. The evaluation uses mean prediction accuracy (MPA), which is calculated by comparing the predicted stock prices to the actual prices across all stocks. The results are based on multiple runs (451 in total) to account for variations in individual stock patterns, achieving an average prediction accuracy that was 0.32% higher than that of the standard LSTM model that uses news data, indicating that the added DP measures not only protect data but also improve robustness and accuracy. Furthermore, DP-LSTM consistently outperformed the standard LSTM across different time durations and settings.[51]

The next approach, Private Aggregation of Teacher Ensembles (PATE), protects sensitive data by splitting the private dataset into separate parts and training a different teacher model on each part. Each teacher then makes predictions on public data, and these predictions are aggregated—while adding a controlled amount of noise—to create labels for training a final "student" model. Since the student model is only trained

DOI: 10.36838/v7i10.18

on the aggregated and noisy outputs rather than directly on private data, it becomes much harder for an attacker to recover any sensitive information. In the results for this model, the researchers revealed that fewer teacher models (N=10) led to a more consistently high accuracy in the student model. However, as N increases, training becomes less stable due to the reduced number of samples per teacher. Additionally, when the privacy parameter (ε) is set lower (indicating stronger privacy), the student model's accuracy is higher (up to 91.92%), but as ε increases (weaker privacy), accuracy drops sharply, highlighting a clear trade-off between privacy and accuracy. When ε is below 14, the model becomes nearly impracticable, showing that balancing privacy and performance in the PATE framework is quite challenging.

Finally, DP-SGD is examined, which has already been explained in section 4.2. This approach evaluated two different gradient clipping sizes (1 and 1.5), revealing that a clipping size of 1.5 results in more stable accuracy across different privacy levels. With DP-SGD, even when the privacy parameter is set to a low value (indicating strong privacy), the models maintain an accuracy of 88%, which is close to the non-private baseline. This performance demonstrates that DP-SGD can effectively balance privacy and utility, outperforming the PATE approach. The differences in noise scales between these methods contribute significantly to their varying performance, with DP-SGD showing better overall stability and higher accuracy under tight privacy requirements.[52]

Federated Learning

Federated Learning (FL) is a machine learning method that lets multiple clients train a shared global model without sending their raw data to a central server. Instead, each client trains its own model using its local data and then sends only the updated model parameters (such as weights or gradients) to the server. The server then combines these updates to improve the global model, ensuring that sensitive data never leaves the client devices (Figure 3).[53]

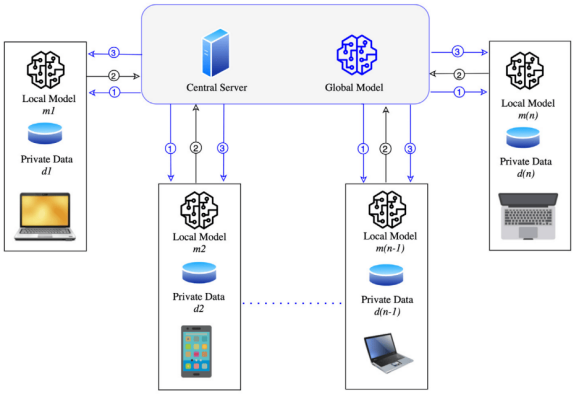


Figure 3: Federated Learning system design: (1) the server initializes and broadcasts a global model to all clients, (2) each client trains the model locally on its private data and sends only model updates back, (3) the server aggregates these updates to refine the global model, which is then redistributed for the next training round—this cycle repeats until convergence.

In FL, a global model is considered valid if it performs better than at least one of the individual local models. Another measure, δ-Accuracy, compares the global model's performance to that of a model trained on the combined dataset, with a small difference indicating good performance. The FL training process generally involves three key steps: the server first initializes the model and distributes it to selected clients; the clients then update their local models using their own data; and finally, the server aggregates these updates to form a new global model. The FedAvg algorithm is a well-known example of this approach, where the server averages the model updates from multiple clients to improve the overall model accuracy while keeping each client's data private.[54]

FL comes in various forms, each for a specific type of data distribution. Vertical Federated Learning is used when different organizations have different sets of features or information on the same individuals. For example, if one company has healthcare information and another one has financial information on the same set of individuals, they can collaborate without ever sharing their raw data. Horizontal Federated Learning is when all participants have the same information on different individuals. A common example is mobile devices, where each phone collects similar data (i.e., keyboard usage) from its various users. Federated Transfer Learning is used when participating entities have different feature spaces. This method helps them transfer knowledge from one domain to another, effectively introducing new data to a model. Cross-silo Federated Learning is when a small number of organizations (silos) consistently participate in the training process. This is typically used by healthcare or financial institutions that have access to rich datasets. Finally, Cross-Device Federated Learning is when a large number of devices (i.e., phones) train a model together. Since such devices often have limited resources, techniques like client selection are used to ensure an efficient training process.[55]

Federated Learning in LLMs:

In this subsection, we will discuss two approaches of FL applied in LLMs in order to alleviate the arising privacy risks.

FATE-LLM is an FL system designed for training LLMs securely and efficiently in a distributed environment. The system consists of three main parts: the Communication-Efficient Hub, the FedLLM Model Hub, and the FedLLM Privacy Hub. The Communication-Efficient Hub incorporates various techniques to reduce the amount of data exchanged between clients and the server. The FedLLM Model Hub gathers different popular LLMs so that they can be used for various applications. The FedLLM Privacy Hub adds extra layers of security to ensure that sensitive data remains private during training. FATE-LLM also supports different training scenarios. Through their experiments on fine-tuning models, both LoRA Federated and P-Tuning-v2 Federated methods perform better than when these methods are applied separately on individual clients. However, the federated approaches fall short on all performance metrics, indicating that while FL improves client performance, there is still significant room for improvement to match the efficiency of centralized methods.[56]

The second approach of FL in LLMs is FEDBiOT, which compresses a pre-trained LLM by splitting it into two parts: an adapter and an emulator. First, the bottom few layers of the original LLM (near the output) are extracted as the adapter, which captures the domain-specific features and is the only part that is fine-tuned locally. The rest of the model forms the non-compressed portion, which is then reduced using uniform layer dropout to create the emulator. To ensure that this compressed emulator behaves similarly to the full model, a pre-alignment step is performed using knowledge distillation on a public dataset, where the emulator is tuned to closely mimic the activations and output distribution of the original model. Finally, LoRA (Low-Rank Adaptation) is applied to both the adapter and the emulator to further reduce communication and computation costs during the federated learning process. Experimental results show that FeDBiOT was able to balance privacy and utility well in FL environments, even when strong privacy measures, such as dropout, are applied. Overall, FedBiOT outperforms baseline methods by maintaining more stable and accurate predictions across different privacy settings. Adjusting the size of the adapter further improves performance, indicating that a larger, more flexible adapter can capture more knowledge from the data while still meeting privacy requirements.[57]

Federated Learning in Transformers:

In this subsection, we will examine two approaches to Federated Learning applied to Transformers.

FedYolo is a multitask FL algorithm that is designed to reduce communication costs and improve robustness when training transformer models. Traditional FedAvg requires sending the full model parameters back and forth, which can be slow and sensitive to differences between clients. In contrast, FedYolo uses a frozen pre-trained transformer (PTF) as a backbone that is loaded once by each client, and then each client only updates small, task-specific modules for their particular tasks. This means that instead of sharing the entire model, clients only send these smaller modules during training, which lowers communication costs. To address potential privacy issues (since the server might learn which client is working on which task), the approach can be modified so that clients send a sparse vector representing their modules, and secure aggregation techniques are used to keep individual contributions private. Experimental results show that FedYolo consistently outperformed the conventional FedAvg method, especially when there were more tasks or when using smaller pre-trained transformer models. When additional local training (personalization) was applied after the federated training, FedYolo still achieved higher accuracy compared to both FedAvg and models that were trained only locally. The performance difference between FedYolo and the other methods became smaller, though, as the size of the PTF increased, which suggests that larger models cannot balance local and federated training as effectively as smaller models.[58]

The second approach is FedPerfix, which is designed to balance global and local learning by only updating the sensitive parts of the model while keeping a frozen PTF backbone. This design helps reduce communication costs and resource usage while still capturing task-specific details from local data, making it especially effective in scenarios with diverse data. The experimental results demonstrate that FedPerfix consistently outperforms the baseline methods, achieving the highest average Top-1 accuracy across the three datasets tested. This approach shows notable improvements in performance, particularly in challenging settings where data distributions are skewed. Client-wise performance analysis reveals that FedPerfix delivers significant accuracy gains—up to a 30% improvement for some clients, indicating huge benefits for most participants. Additionally, a resource analysis shows that FedPerfix significantly reduces communication costs compared to other methods, even though it requires slightly more storage and computation.[59]

Federated Learning in Marketing:

Privacy is a cornerstone in marketing because if customers lose their trust in a company, that is a potential revenue loss, and with the rise of LLMs and Transformers, this risk has only been increasing. In this subsection, we will analyze an approach to FL in this industry.

The AROUND system is designed as a social network focused on small- and medium-sized businesses, using an intelligent advertising platform to deliver personalized shopping experiences. The system divides its users into two groups: traders (business owners) and customers. The traders use physical beacons that are placed in various locations to send location-based notifications to nearby customers. When a customer comes within range, their app connects to a central cloud server, which then collects their current location, profile, and even their selected daily mood. Using this information, the system employs a clustering model to determine the most appropriate advertisement for that customer based on past behavior and preferences. This setup allows for localized and personalized ad recommendations while maintaining a streamlined and efficient communication process between the devices, beacons, and cloud. The experimental results reveal several key findings about this approach. Firstly, as the number of devices increases from 25 to 100, the average device usage time drops significantly (a 64% decrease). This shows that the architecture effectively distributes the workload. Second, while FL does result in slightly higher communication overhead compared to a centralized approach, this increase is manageable for this approach as it scales well with the number of users. Third, the recommender system's performance improved remarkably, with both higher F1 scores and click-through rates (CTR). For instance, the FL approach achieves an average CTR of 53.9%, outperforming both central ML and random ad selections. Additionally, user engagement metrics such as the time spent on ads and interaction rates are also notably better in the FL mode. Finally, the opt-out rates are lower, indicating that personalized ads are more effective.[60]

Federated Learning in Healthcare:

Next, we will examine a Federated Learning approach that has been applied to Healthcare, along with some real-world applications that have already occurred in places like the U.S. and the U.K.

DOI: 10.36838/v7i10.18

In the first approach, each hospital or lab (the data owner) keeps its own electronic health records (EHRs) in a private format while training a local model on its data. These local models, which might include various types of medical information like imaging results, diagnostic reports, and prescriptions, are then combined by a central or decentralized manager using secure aggregation techniques. In addition, the system includes modules to process and convert the data into a consistent format, verifying that learning algorithms do not leak private information, and enforcing strict access controls and encryption. This ensures that even as data is used for analysis, patient data remains secure. Their results indicate that the proposed setup effectively maintains data privacy while still allowing for meaningful and accurate analysis across multiple institutions. It highlights that by keeping data local and only sharing aggregated model updates, the architecture minimizes the risk of data breaches, supports robust analysis, and meets regulatory requirements.[61]

In one project in the US, a group of institutions, including the University of Pennsylvania, used powerful Intel processors and special security tools (Intel SGX) to run FL for image analysis. Hospitals run the FL tasks locally on their own data and then share only the updated model information with a central server. In another project in the UK, Nvidia partnered with King's College London and Owkin to build an FL platform for the National Health Service. This setup uses Nvidia's technology to allow hospitals to collaborate on important medical problems like cancer and heart disease while keeping their patient data private. There was also an international FL project done for COVID-19, where hospitals from China, Italy, and Japan combined their CT scan data to create models that better segment COVID-19 regions in chest scans. By pooling data in this way, the FL method helps overcome the lack of data available at individual hospitals while maintaining strict privacy.[62]

Federated Learning in Finance:

Finally, we will take a look at one last FL approach applied to the finance industry. By combining SMPC with FL, financial institutions can collaboratively train robust models for tasks like fraud detection and credit risk assessment without transferring sensitive customer data, thereby simplifying compliance with regulations and enhancing predictive performance across diverse datasets. In this approach, each client first computes its own model weights from its local data. Then, each client masks its weight by adding and subtracting random values before sending it to the central server. This masking ensures that the individual weights are hidden. The server then aggregates these masked weights—typically by taking a weighted average—resulting in a global model that is equivalent to what would have been obtained without masking/encryption, but without exposing any client's private data. Furthermore, the protocol is designed to work iteratively, where the clients only need to set up the secure randomness once at the beginning, and then update it in each round using a pseudorandom generator.[63]

■ Discussion

Analysis:

Table 2: This table contrasts three key PPMs. Homomorphic Encryption delivers the strongest confidentiality by keeping data encrypted at all times, but at the cost of heavy computation and storage overhead. Differential Privacy offers a configurable privacy–utility balance by adding noise, which can modestly reduce accuracy. Federated Learning ensures data never leaves devices, enabling collaborative model training with moderate communication burdens and challenges from non-IID data.

	Privacy	Utility	Summary
Homomorphic Encryption	Very high—data remains encrypted throughout processing[17]	Substantial computational slowdown and larger storage[18]	Provides the strongest confidentiality guarantees but incurs significant performance and resource costs
Differential Privacy	Adjustable—added noise obscures individual contributions [27]	Reduced accuracy to some degree[28]	Offers a tunable balance between privacy and model quality by introducing calibrated noise
Federated Learning	Strong—raw data never leaves local devices[42]	Moderate overhead from communication and uneven data distributions[43]	Enables collaborative training without sharing raw data but must manage communication and data-heterogeneity challenges

Homomorphic Encryption (HE) Limitations

HE enables computation on encrypted data, but faces several bottlenecks:

- Computational Overhead: Encrypted operations can be tens to thousands of times slower than plaintext equivalents, making HE impractical for large-scale or real-time applications (see Sections 3.4, 3.5, 3.6).
- Ciphertext Expansion: Encrypting data typically inflates ciphertext size, leading to higher storage and transmission costs—another barrier for resource-constrained environments.
- Noise Management: Each homomorphic operation introduces noise; fully homomorphic schemes require expensive bootstrapping to control this noise, further degrading performance. Without bootstrapping, HE schemes can support only a limited number of operations.
- Engineering Complexity: Efficient, secure HE (especially FHE) implementations are intricate and demand significant engineering effort. Deploying these schemes at scale remains challenging, though Sections 3.2 and 3.3 demonstrate that smaller models (e.g., BERTtiny) can successfully leverage HE without sacrificing utility.

Differential Privacy (DP) Limitations

DP protects individual data points by adding noise to query results, but entails trade-offs:

- Privacy–Utility Trade-off: Stronger privacy (lower ε) requires more noise, which can degrade model accuracy. Conversely, reducing noise (higher ε) weakens privacy guarantees.
- Cumulative Loss: Every query on the same dataset consumes part of the privacy budget; repeated queries can eventually erode the overall privacy guarantee if not carefully managed.
- Integration Overhead: Incorporating DP into complex ML pipelines—particularly deep learning—adds computational cost and necessitates expertise to tune ε and stabilize training. Nevertheless, techniques like Re-Attention (Section 4.4) and DP-SGD variants (Section 4.6) show that practical DP implementations are achievable without crippling utility.

Federated Learning (FL) Limitations

FL lets multiple clients train a shared model without centralizing raw data, but introduces its own challenges:

- Data Heterogeneity: Clients often hold non-IID data, leading to convergence issues and inconsistent model performance across nodes.
- System Heterogeneity: Differences in client compute power, network quality, and data volumes can hinder synchronization and slow down training.
- Communication Overhead: Frequent exchange of model updates between many clients and a central server generates substantial bandwidth consumption, especially as participant count grows.
- Residual Privacy Risks: Although raw data stays local, model updates themselves can leak sensitive information if adversaries launch inference attacks.
- Participation and Incentives: Ensuring enough clients contribute high-quality data is difficult—clients may lack motivation, resources, or trust in the process. Despite these issues, Sections 5.2–5.6 survey FL approaches that mitigate such concerns, and real-world healthcare deployments (Section 5.5) illustrate practical FL success.

Overall Comparison

Each mechanism—HE, DP, and FL—offers unique strengths and faces distinct limitations. HE delivers the strongest confidentiality, but at high computational and engineering costs. DP provides quantifiable privacy guarantees but can reduce data utility and requires careful budget management. FL allows collaborative training without central data pooling but must contend with heterogeneity, communication overhead, and residual privacy risks. Recent advances across all three fields show promising directions, yet no single approach fully addresses end-to-end privacy, utility, and scalability requirements for high-stake applications (Table 2).

Further Research Pathways

We propose a comprehensive solution that combines Homomorphic Encryption, Differential Privacy, and Federated Learning in a complementary way. During the training phase, we propose using an FL base where each client trains its local model. However, each local model would also be using DP techniques, ensuring that sensitive data remains protected by adding controlled noise. These locally trained models then share their updates with a central server through federated learning, where the server aggregates the model parameters to create a global model, all without ever accessing the raw data. In practice, this requires careful tuning of the DP noise budget at each client to avoid excessive utility loss, as well as ensuring that client devices have sufficient CPU/GPU resources to handle on-device DP mechanisms. Finally, at the end of the training process—when model parameters need to be transmitted and potentially used in downstream applications—homomorphic encryption is applied. This final step secures the transmission of model parameters, ensuring that any sensitive information is protected even during communication. Implementing HE at this stage introduces additional computational and latency overhead, particularly for larger model sizes, necessitating optimized key-management schemes and potential batching strategies. By integrating these methods, the proposed approach could potentially achieve a stronger balance between privacy guarantees and model utility, making it a promising solution for sensitive domains such as healthcare, finance, and marketing.

In a real-world healthcare application, imagine that several hospitals want to work together to develop a predictive model for patient readmission risk without sharing sensitive patient details (Figure 4). Using this proposed system design, each hospital would train its own local model on its own private data while also using differential privacy techniques, meaning that the patient data would be perturbed with controlled noise so that individual records cannot be reidentified. The local models would then be combined using federated learning techniques, where the hospitals would only send their encrypted model parameters to a central server, which would then aggregate them to form a global model. However, in a large federation of hospitals, data heterogeneity (non-IID distributions) and intermittent network connectivity can challenge convergence; robust aggregation methods and mechanisms for handling straggler clients are required to mitigate these issues. Finally, once this training is complete, homomorphic encryption techniques will be applied when transmitting and using the global model parameters, ensuring that even during communication, data remains secure and private. Nevertheless, end-to-end HE can impose significant bandwidth costs and decryption delays—especially if real-time inference is needed—so practical deployments must weigh encryption parameters against acceptable latency. This system would allow hospitals to develop a robust predictive model without ever needing to share their raw sensitive data, ensuring privacy from the start to the end.

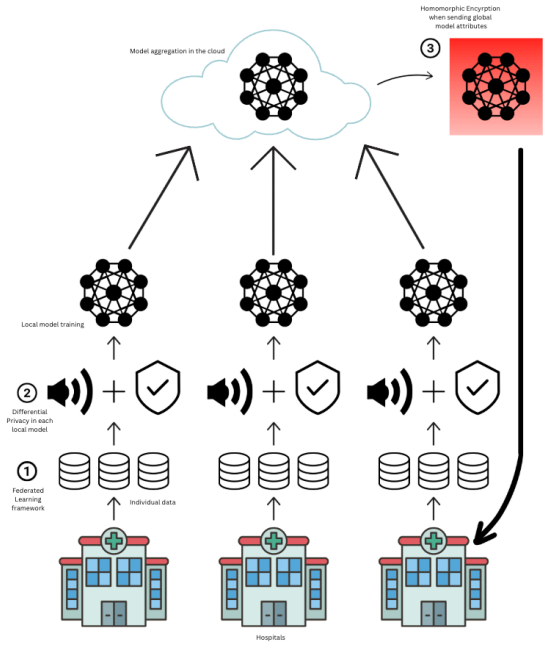


Figure 4: Three-step integrated privacy pipeline where clients train locally with Differential Privacy, share only noisy updates via Federated Learning for secure aggregation, and then encrypt the final model parameters with Homomorphic Encryption before any transmission.

DOI: 10.36838/v7i10.18

Conclusion

In conclusion, this paper has examined how three key PPMs–Homomorphic Encryption, Differential Privacy, and Federated Learning–are being applied to Large-Language Models and Transformers in the marketing, healthcare, and finance industries. The reviewed approaches demonstrate promising capabilities in safeguarding sensitive data while maintaining high model performance, yet they also highlight ongoing challenges such as trade-offs between privacy and utility, computational overhead, and scalability concerns. Policy frameworks should be updated to recognize and incentivize the adoption of PPMs, providing clear guidelines on acceptable privacy thresholds and compliance requirements. Industry practitioners are encouraged to establish cross-functional teams that include legal, technical, and operational stakeholders to streamline implementation and ensure alignment with regulatory mandates. Moving forward, further research is needed to refine these methods, explore new applications, and develop robust frameworks that can be seamlessly deployed in real-world scenarios. Ultimately, our findings underscore the importance of continuing to innovate in the field of privacy-preserving machine learning to meet the evolving demands of secure and ethical data utilization. Practitioners should pilot integrated PPM solutions in controlled environments, measure performance against privacy objectives, and share best practices across sectors to accelerate broader adoption.

Acknowledgments

I would like to express my deepest gratitude to Dr. Siddharth Krishnan for his unwavering support, invaluable guidance, and dedicated time throughout this research project.

References

1. Melillio, Pietro. "Massive Data Exposure on X (Twitter) Affects 200 Million Users." *RedHotCyber*, 2024. https://www.redhotcyber.com/en/post/massive-data-exposure-on-x-twitter-affects-200-million-users/.
2. Alder, Steve. "The Biggest Healthcare Data Breaches of 2024." *TheHIPAAJournal*, 2025. https://www.hipaajournal.com/biggest-healthcare-data-breaches-2024/#:~:text=The%20largest%20healthcare%20data%20breach,paid%20a%20$22%20million%20ransom.
3. Mascellino, Alessandro. "Cyber-Attack on Evolve Bank Exposed Data of 7.6 Million Customers." *InfoSecurity Magazine*, 2024. https://www.infosecurity-magazine.com/news/cyber-attack-evolve-bank-exposed/.
4. "Key Issues." *General Data Protection Regulation* (*GDPR*), 9 Oct. 2018, gdpr-info.eu/issues/.
5. Yang, *et. al.* "Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond." *ACM Digital Library*, 2024. https://dl.acm.org/doi/10.1145/3649506#sec-5-3.
6. Carlini, *et. al.* "The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks." *USENIX SECURITY SYMPOSIUM*, 2019. https://www.usenix.org/conference/usenixsecurity19/presentation/carlini.
7. Chen, *et. al.* "A Survey on Large Language Models for Critical Societal Domains: Finance, Healthcare, and Law." *arXiv*, 2024. https://arxiv.org/abs/2405.01769.
8. Kandpal, *et. al.* "Deduplicating Training Data Mitigates Privacy Risks in Language Models." *arXiv*, 2022. https://arxiv.org/abs/2202.06539.
9. Okta. "Privacy vs. Security: Exploring the Differences & Relationship." *Okta*, 2024. https://www.okta.com/identity-101/privacy-vs-security/.
10. Solove, J. Daniel. "Understanding Privacy." *Google Book*s, 2008. https://books.google.com/books?hl=en&lr=&id=eSrnEAAAQBAJ&oi=fnd&pg=PT7&ots=MdrSiI0z2i&sig=r6-DSqav2BulyCVMHwEaHayrHOk#v=onepage&q&f=false.
11. Rosenbush, Steven. "Large Language Models Pose Growing Security Risks." *Mint*, 21 Feb. 2025, www.livemint.com/technology/large-language-models-pose-growing-security-risks-11740100764542.html.
12. Unknown. "How does cybersecurity affect marketers? 5 tips to keep your brand safe and strong." *Exclaimer*, 2024. https://exclaimer.com/blog/how-does-cybersecurity-affect-marketers/.
13. Kshetri, Nir. "Generative artificial intelligence in marketing: Applications, opportunities, challenges, and research agenda." *ScienceDirect*, 2024. https://www.sciencedirect.com/science/article/abs/pii/S026840122300097X.
14. Cunha, *et. al.* "A survey of privacy-preserving mechanisms for heterogenous data types." *ScienceDirect*, 2021. https://www.sciencedirect.com/science/article/pii/S1574013721000435.
15. IBM. "What is cryptography?" *IBM*. https://www.ibm.com/think/topics/cryptography.
16. Devane, Heather. "Everything You Need to Know About K-Anonymity." *Immuta*, 2021. https://www.immuta.com/blog/k-anonymity-everything-you-need-to-know-2021-guide/.
17. Niculaescu, Oana. "K-Anonymity and cluster based methods for privacy." *GitHub*, 2017. https://elf11.github.io/2017/04/22/kanonymity.html.
18. Olzak, Tom. "Homomorphic Encryption: How It Changes the Way We Protect Data." *SpiceWorks*, 2022. https://www.spiceworks.com/it-security/data-security/articles/how-homomorphic-encryption-protects-data/.
19. Ogburn, *et. al.* "Homomormphic Encryption." *ScienceDirect*, 2013. https://www.sciencedirect.com/science/article/pii/S1877050913011101.
20. *Can Homomorphic Encryption Be Practical? | Proceedings of the 3rd ACM Workshop on Cloud Computing Security Workshop*, dl.acm.org/doi/abs/10.1145/2046660.2046682. Accessed 10 June 2025.
21. Brown University. "Applying Fully Homormophic Encryption (Part I)." *Brown University*, 2012. https://esl.cs.brown.edu/blog/applying-fully-homomorphic-encryption-part-1/.
22. Geelen, *et. al.* "BASALISC: Programmable Hardware Accelarator for BGV Fully Homomorphic Encryption." *arXiv,* 2023. https://esl.cs.brown.edu/blog/applying-fully-homomorphic-encryption-part-1/.
23. *Applications of Homomorphic Encryption*, www.researchgate.net/profile/Roger_Hallman2/publication/320976976_APPLICATIONS_OF_HOMOMORPHIC_ENCRYPTION/links/5a051f4ca6fdcceda0303e3f/APPLICATIONS-OF-HOMOMORPHIC-ENCRYPTION.pdf. Accessed 10 June 2025.
24. *Encryption-Friendly LLM Architecture*, arxiv.org/pdf/2410.02486. Accessed 10 June 2025.
25. *A Fully Homomorphic Encryption Scheme a ...*, crypto.stanford.edu/craig/craig-thesis.pdf. Accessed 10 June 2025.
26. Raeini, *et. al* "Privacy-Preserving Large-Language Models (PPLLMs)." *SSRN*, 2023. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4512071.
27. Macdonald-Ross, Michael, and Robert Waller. "The Transformer Revisited." *Information Design Journal*, John Benjamins, 1 Jan. 1998, www.jbe-platform.com/content/journals/10.1075/idj.9.2-3.06mac.

28. Zimerman, *et. al*. "Converting Transformers to Polynomial Form for Secure Inference Over Homomorphic Encryption." arXiv, 2023. https://arxiv.org/abs/2311.08610.
29. Rovida, *et. al*. "Transfomer-based Language Models and Homomorphic Encryption: An Intersection with BERT-tiny." *ACM*, Digital Library, 2024. https://dl.acm.org/doi/abs/10.1145/3643651.3659893.
30. Mittal, Rohit. "Improving Inference Privacy for Large Language Models using Fully Homomorphic Encryption." *University of California, Berkeley*, 2024. https://www2.eecs.berkeley.edu/Pubs/TechRpts/2024/EECS-2024-225.pdf.
31. *Principles of Management - Cloudfront*.Net, d3bxy9euw4e147.cloudfront.net/oscms-prodcms/media/documents/PrinciplesofManagement-OP.pdf. Accessed 10 June 2025.
32. Cui, *et. al.* "Digital marketing program design based on abnormal consumer behavior data classification and improved homomorphic encryption algorithm." *PeerJ*, 2023. https://peerj.com/articles/cs-1690/.
33. Dash, Sabyasachi, *et al*. "Big Data in Healthcare: Management, Analysis and Future Prospects - Journal of Big Data." *SpringerLink*, Springer International Publishing, 19 June 2019, link.springer.com/article/10.1186/s40537-019-0217-0?utm_campaign=scm&campaignid=70161000001CouwAAC%3FStartPage.
34. *Influence of the HIPAA Privacy Rule on Health Research | Law and Medicine | JAMA | Jama Network,* jamanetwork.com/journals/jama/fullarticle/209447. Accessed 10 June 2025.
35. Roumpies, *et. al*. "A Review of Homomorphic Encryption and its Contribution to the Sector of Health Services." *ACMDL*, 2024. https://dl.acm.org/doi/pdf/10.1145/3635059.3635096.
36. Avanidhar, Subrahmanyam. "Big Data in Finance: Evidence and Challenges." *Borsa Istanbul Review*, Elsevier, 8 Aug. 2019, www.sciencedirect.com/science/article/pii/S2214845019302650.
37. Peng, *et. al.* "Homomorphic Encryption Application on FinancialCloud Framework." *IEEE Xplore*, 2016. https://ieeexplore.ieee.org/abstract/document/7850013.
38. Alizadeh, Essi. "The ABCs of Differential Privacy." *Towards Data Science*, 2023. https://towardsdatascience.com/abcs-of-differential-privacy-8dc709a3a6b3/.
39. "What is Differential Privacy?" *IEEE Digital Privacy*. https://digitalprivacy.ieee.org/publications/topics/what-is-differential-privacy.
40. Stalla-Bourdillon, Sophie. "What is Differential Privacy? A Guide for Data Teams." *IMMUTA*, 2024. https://www.immuta.com/blog/why-differential-privacy-should-be-top-of-mind-for-data-science-and-governance-teams/.
41. Martineau, Kim. "What is federated learning?" *IBM*, 2022. https://research.ibm.com/blog/what-is-federated-learning.
42. Salim, *et. al*. "A Privacy-Preserving Local Differential Privacy-Based Federated Learning Model to Secure LLM from Adversarial Attacks." *Human-centric Computing and Information Sciences*, 2024. https://hcisj.com/articles/issue_view.php?wr_id=533&page=&ckattempt=2.
43. Behnia, *et. al*. "EW-Tune: A Framework for Privately Fine-Tuning Large Language Models with Differential Privacy." *IEEE Xplore*, 2022. https://ieeexplore.ieee.org/abstract/document/10031034.
44. Singh, *et. al*. "Whispered Tuning: Data Privacy Preservation in Fine-Tuning LLMs through Differential Privacy." *Scientific Research*, 2024. https://www.scirp.org/journal/paperinformation?paperid=130659.
45. Charles, *et. al*. "Fine-Tuning Large Language Models with User-Level Differential Privacy." *arXiv*, 2024. https://arxiv.org/abs/2407.07737.
46. Ding, *et. al*. "Delving into Differentially Private Transformer." *arXiv*, 2024. https://arxiv.org/abs/2405.18194.
47. Ponomareva, *et. al*. "Training Text-to-Text Transformers with Privacy Guarantees." *ACL Anthology*, 2022. https://aclanthology.org/2022.findings-acl.171/.
48. Brown, *et. al.* "Exploring Differentially Privacy Approaches for Bid Optimization in Digital Marketing." *ResearchGate*, 2024. https://www.researchgate.net/profile/Sophia-Lee-32/publication/387946664_Exploring_Differential_Privacy_Approaches_for_Bid_Optimization_in_Digital_Marketing/links/67837b9f8210a977a183ac2f/Exploring-Differential-Privacy-Approaches-for-Bid-Optimization-in-Digital-Marketing.pdf.
49. Dankar, *et. al.* "Practicing Differential Privacy in Health Care: A Review." *Microsoft Word,* 2013. https://www.tdp.cat/issues11/tdp.a129a13.pdf.
50. Choudhury, *et. al*. "Differential Privacy-enabled Federated Learning for Sensitive Health Data." *arXiv*, 2020. https://arxiv.org/pdf/1910.02578.
51. Li, *et. al.* "DP-LSTM: Differential Privacy-inspired LSTM for Stock Prediction Using Financial News." *arXiv*, 2019. https://arxiv.org/abs/1912.10806.
52. Chen, *et. al*. "Financial Vision Based Differentially Private Applications." *arXiv*, 2021. https://arxiv.org/abs/1912.10806.
53. Mothukuri, Viraaji, *et al*. "A Survey on Security and Privacy of Federated Learning." *Future Generation Computer Systems,* North-Holland, 10 Oct. 2020, www.sciencedirect.com/science/article/abs/pii/S0167739X20329848.
54. Yin, *et. al*. "A Comprehensive Survey of Privacy-preserving Federated Learning: A Taxonomy, Review, and Future Directions." *ACMDL*, 2021. https://dl.acm.org/doi/10.1145/3460427.
55. Mammen, Priyanka Mary. "Federated Learning: Opportunities and Challenges." *arXiv,* 2021. https://arxiv.org/pdf/2101.05428/.
56. Fan, *et. al.* "FATE-LLM: An Industrial Grade Federated Learning Framework for Large Language Models." *arXiv*, 2023. https://arxiv.org/pdf/2310.10049.
57. Wu, *et. al*. "FedBiOT: LLM Local Fine-tuning in Federated Learning without Full Model." *ACMDL*, 2024. https://dl.acm.org/doi/pdf/10.1145/3637528.3671897.
58.Zhang, *et. al.* "FedYolo: Augmenting Federated Learning with Pretrained Transformers." *arXiv*, 2023. https://arxiv.org/pdf/2307.04905.
59. Sun, *et. al*. "FedPerfix: Towards Partial Model Personalization of Vision Transformers in Federated Learning." *arXiv*, 2023. https://arxiv.org/pdf/2308.09160.
60. Seyghaly, *et. al.* "A Comprehensive Architecture for Federated Learning-Based Smart Advertising." *MDPI*, 2024. https://www.mdpi.com/1424-8220/24/12/3765.
61. Antunes, *et. al*. "Federated Learning for Healthcare: Systematic Review and Architecture Proposal." *ACMDL*, 2022. https://dl.acm.org/doi/abs/10.1145/3501813.
62.Nguyen, *et. al*. "Federated Learning for Smart Healthcare: A Survey." *ACMDL*, 2022. https://dl.acm.org/doi/abs/10.1145/3501296.
63. Byrd, *et. al*. "Differentially private secure multi-party computation for federated learning in financial applications." *ACMDL*, 2021. https://dl.acm.org/doi/abs/10.1145/3383455.3422562.

Author

Avani is a 11th grader at Saint Francis High School in California with a strong passion for security in Generative AI. She is also intrigued by the intersection of technology and management, and she hopes to explore how AI systems can be integrated into business strategies while maintaining robust privacy.

DOI: 10.36838/v7i10.18

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Comparative Analysis of Ascorbic Acid Content in Fruits: Impact of Temperature and Storage Conditions

Dhruv S. Prashant

Oberoi International School, JVLR, Jogeshwari East, Mumbai, 400060 India; dhruv.s.prashant@gmail.com

ABSTRACT: Ascorbic Acid (Vitamin C) is an indispensable nutrient for its crucial role in immune functions, as well as a multitude of other health benefits. This study aims to compare the ascorbic acid content of 3 fruits (orange, guava, lemon) using a redox titration method using Iodine. Among the fruits analyzed, guava contained the highest amount of Vitamin C (307 mg/ 100 g of fruit sample), followed by orange (142 mg/ 100 g) and lemon (103 mg/ 100 g). All three fruits exhibited a reduction in ascorbic acid content after boiling for 3 minutes. The observed reduction in ascorbic acid concentration was 8% for guava, 13% for orange, and 4.6% for lemon. Additionally, storing fruits for 3 days in an open container at an ambient temperature of 28 °C also led to degradation with a reduction of 8% for guava, 20.7% for orange, and 6.6% for lemon. The present study suggests that storage conditions and processing methods impact the retention of ascorbic acid, and the extent of reduction differs depending on the type of fruit.

KEYWORDS: Biochemistry, General Biochemistry, Ascorbic Acid, Fruits, Temperature.

Introduction

Vitamins are a group of organic compounds that are needed for the healthy functioning of the body. Vitamins are categorized as fat-soluble (Vitamins A, D, E, and K) and water-soluble (Vitamins B and C). L-Ascorbic acid (Vitamin C) is an important vitamin essential for various metabolic processes and is primarily considered a potent antioxidant. Its physiological role is extensive and includes various processes such as the absorption of iron as a key facilitator, hormone synthesis, and playing significant roles in wound healing and collagen production. Humans cannot synthesize Vitamin C and hence are fully dependent on its intake from diet.[1] Citrus fruits, which belong to the genus Citrus of the family Rutaceae, such as oranges, lemons, are well established sources of Vitamin C.[2] Apart from Citrus fruits, there are fruits like guava, which also have significant amounts of Vitamin C. Guava belongs to the genus Psidium guajava in the family Myrtaceae. There have been extensive studies on the Vitamin C content in guava, and it has been established as a potent source of Vitamin C.[3-5] In addition, commercial fruit juices contain added Vitamin C to enhance nutritional value.

The stability of Vitamin C is dependent on several factors, including exposure to oxygen, temperature, light, and storage conditions. These parameters can affect Vitamin C concentration during processing and storage.

Given the vital role of this vitamin in metabolism and its exclusive dietary origin, this study aims to quantitatively compare the levels of ascorbic acid in commonly consumed fruits (oranges, lemons, and guavas). Further, the study estimates the impact of heat treatment and ambient storage duration on the degradation of ascorbic acid. This study will provide insight into how processing and storage conditions influence nutrient conservation.

Methods

Preparation of fresh fruit samples:

Mandarin Oranges, Thai Pink Guava, and Indian Lemon were procured from an online vendor. Details of the procurement are as below.

Table 1: Details of the sourcing of fruit samples for the experiments.

Fruit Name	Name of Vendor	Variety/ Scientific Name	Extent of Ripeness
Thai Pink Guava	Bhagavati Stores Pvt Ltd, JVLR, Mumbai	*Psidium guajava,* Country of Origin: India	Fully ripe
Mandarin Orange	Bhagavati Stores Pvt Ltd, JVLR, Mumbai	*Citrus reticulata,* Country of Origin: India	Fully ripe
Indian Lemon	Bhagavati Stores Pvt Ltd, JVLR, Mumbai	*Citrus limon,* Country of Origin: India	Fully ripe

100 g sample of 3 fruits (Mandarin Oranges, Thai Guava, or Lemon) was cut into small pieces (without the peel for orange and lemon) and blended in an electric blender. 10 ml portions of distilled water were repeatedly added multiple times, and the liquid was subsequently carefully decanted into a conical flask. Finally, the blended fruit juice was strained through a cheesecloth, rinsing the pulp with 10 ml portions of water. The extracted solution was made up to 100 mL with distilled water.

Preparation of the standard iodine solution:

Analytical-grade Potassium Iodide was procured from Nice Chemicals (Pvt) Ltd. Analytical-grade Iodine crystals (I2) were procured from Research-Lab Fine Chem Industries. 2 g of Potassium iodide was added to a 100 ml beaker. 1.3 g of iodine crystals were weighed on an electronic weighing scale and added to the same beaker, and swirled until the iodine crystals were fully dissolved. The solution was transferred to a 1-liter

DOI: 10.36838/v7i10.31

volumetric flask, and the resulting solution was made up to the 1-liter mark with distilled water.

Preparation of the starch indicator:

Soluble Starch was procured from Spectrum Reagents and Chemicals Pvt Ltd. 0.25g of soluble starch was added to 50 ml of near-boiling water in a 100 ml conical flask and stirred until it was fully dissolved.

Redox titration using iodine solution:

The method for determining the ascorbic acid (Vitamin C) content in solution uses redox titration using iodine solution (Redox titration method, University of Canterbury). Upon the addition of iodine during titration, the ascorbic acid is oxidized to dehydroascorbic acid, and the iodine is reduced to iodide ions.

Ascorbic acid + $I_2 \rightarrow I^-$ + dehydroascorbic acid

Standard Vitamin C tablets were procured from Abbot Healthcare Limited (Brand Limcee® 500 mg, Batch No. HAJA4137, Expiry date, April 2026). 1 Limcee® (500 mg) Vitamin C Chewable tablet was dissolved in 200 ml of distilled water to make the standard solution. In the first experiment, a 20 ml aliquot of the standard solution (50 mg of Vitamin C from Limcee® 500 mg) was added into a 250 ml conical flask using a pipette, and 150 ml of distilled water was added. 1 ml of starch indicator solution was added. This was titrated with the prepared iodine solution (0.005 mol/ litre) and the end point was identified with the first permanent trace of the dark blue-black colour. The standard titer value (Cstandard) for 50 mg of Vitamin C was determined.

In the second experiment, a 20 ml aliquot of the fruit sample solution was added to a 250 ml conical flask using a pipette, and 150 ml of distilled water was added. 1 ml of starch indicator solution was added. This was titrated with the prepared iodine solution (0.005 mol/ Litre), and the endpoint was identified as the first permanent trace of the dark blue-black colour. This was done for all three fruits (orange, guava, lemon) until the endpoint was reached.

Preparation of samples for high temperature and extended storage:

In the third experiment, the fruit samples were boiled for 3 minutes at 100 degrees, and the titration was carried out as mentioned previously with 20 ml aliquots. In the fourth experiment, the fruit samples were stored in open conical flasks for 3 days, at a temperature of 28 °C, to study the effects of storage on the concentration of ascorbic acid. The titration was carried out as mentioned previously with 20 ml aliquots.

Determination of ascorbic acid concentration:

The titer volume for each sample in all the above experiments was compared against the standard Vitamin C titer value ($C_{standard}$) for determining the concentration of ascorbic acid in the samples (C_{sample})

The calculation of the amount of ascorbic acid in the samples (C_{sample}) was calculated using the formula below.

C_{sample} = (Observed titer value / $C_{standard}$) X (250) mg / 100g of the sample

RESULTS

Ascorbic acid concentration varies in different freshly prepared fruit juices:

Ascorbic acid is a water-soluble micronutrient that is essential for the growth of healthy cells. First, the concentration of ascorbic acid in Limcee® Tablet was estimated, which served as a positive control. Titration results are shown in Figure 1. Titer value was determined to be 24.4 ml. This titer value represents the equivalent of 50 mg of ascorbic acid. Subsequently, the concentration of ascorbic acid in freshly prepared juices of different fruits was estimated by the redox titration method. Titration start and end point images for orange, guava, and lemon are presented in Figures 2, 3, and 4, respectively. The endpoint for the fresh orange sample (Figure 2) was achieved at a titer value of 13.8 mL, translating to ascorbic acid content of 142.07 mg/ 100 g. The endpoint for the fresh guava sample (Figure 3) was achieved at a titer value of 30 ml, translating to ascorbic acid content of 307.3 mg/ 100 g. The endpoint for the fresh lemon sample (Figure 4) was achieved at a titer value of 10.06 ml, translating to ascorbic acid content of 103.14 mg/ 100 g. Figure 5 represents the concentration of ascorbic acid in mg per 100 g of the fruit samples. Table 1 mentions the calculated values of the concentration of ascorbic acid. Among the 3 fruits, guava juice contained the highest levels of ascorbic acid (307.3 mg/ 100 g), followed by orange (142.07 mg/ 100 g) and lemon (103.14 mg/ 100g). This is in line with similar research done earlier.[5]

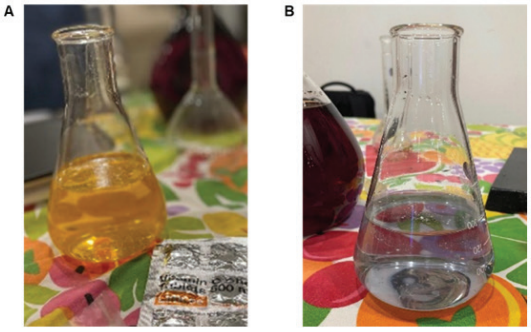


Figure 1: Redox titration results for ascorbic acid standard (Limcee® Tablet): A) Titration start point. B) Titration endpoint. The blue-black color indicates the endpoint of the titration.

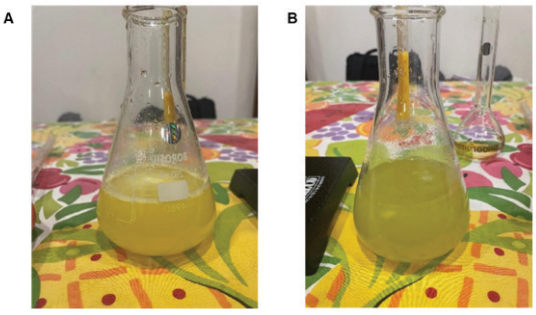


Figure 2: Redox titration results for a freshly prepared orange juice sample: A) Titration start point. B) Titration endpoint. The appearance of the blue-black complex indicates the endpoint of the titration.

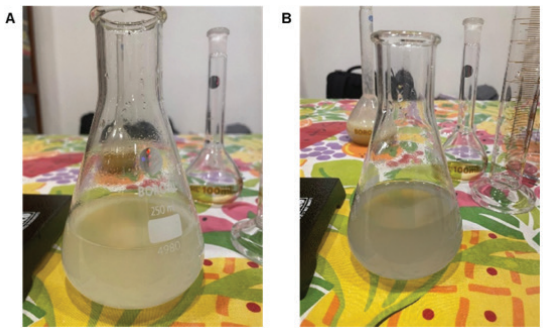


Figure 3: Redox titration results for freshly prepared guava juice sample: A) Titration start point. B) Titration endpoint. The appearance of a blue-black complex indicates the titration endpoint.

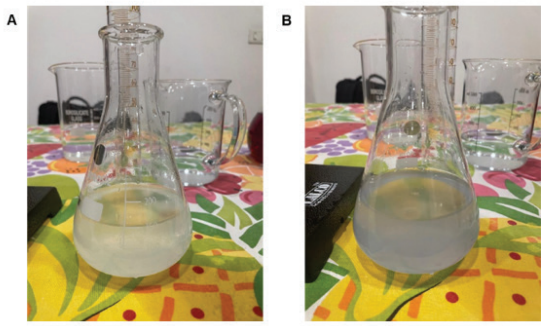


Figure 4: Redox titration results for a freshly prepared lemon juice sample: A) Titration start point. B) Titration endpoint. The appearance of a blue-black complex indicates the endpoint of the titration.

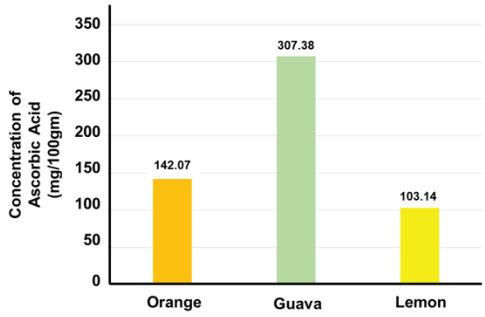


Figure 5: The concentration of ascorbic acid varies in different fruit juice samples. The bar graph represents the concentration of ascorbic acid (mg) per 100 g of fruit sample. The figure indicates the highest amount of Vitamin C in guava, followed by orange, and then lemon.

Table 2: The concentration of ascorbic acid (in mg) in 100 g of freshly prepared fruit samples indicates the highest concentration in guava, followed by orange and lemon.

Fruit Sample	Concentration of ascorbic acid in mg/ 100 g
Orange	142.07 mg/ 100g
Guava	307.38 mg/ 100 g
Lemon	103.14 mg/ 100 g

Boiling fruit samples reduces the ascorbic acid concentration:

Ascorbic acid is a temperature-sensitive vitamin, and an increase in temperature significantly reduces the concentration of ascorbic acid. To study the effect of temperature, separate fresh juice of orange, guava, and lemon, each followed by boiling for 3 minutes, was prepared. Redox titration of both freshly prepared juice and boiled juice was carried out. Boiling of fruit juice samples reduced the concentration of ascorbic acid as compared to freshly prepared fruit juice samples (Figure 6 and Table 2). All the samples showed a degradation of ascorbic acid after boiling for 3 minutes. The concentration of ascorbic acid was estimated to be 123.63 mg/ 100 g for the boiled orange juice sample, 282.79 mg/ 100 g for the boiled guava juice sample, and 98.361 mg/ 100 g for the boiled lemon juice sample. These values indicate a reduction of 13%, 8%, and 4.6% of ascorbic acid content for orange, guava, and lemon, respectively, when subjected to boiling for 3 minutes.

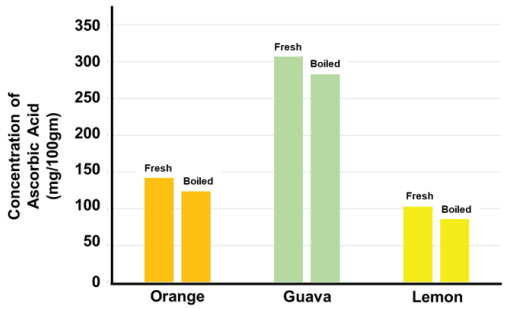


Figure 6: Boiling degrades ascorbic acid concentration in fruit juice samples. The bar graph represents the concentration of ascorbic acid (mg) per 100 g of fruit sample in freshly prepared and boiled fruit juices of orange, guava, and lemon. The figure indicates the reduction in Vitamin C content in the boiled samples across all 3 fruits.

Table 3: Concentration of ascorbic acid in 100 g of freshly prepared fruit juice and 100 g of fruit juice samples boiled at 100 °C for 3 minutes (orange, guava, lemon).

Fruit Sample	Concentration of Ascorbic Acid (mg) per 100 g of sample	
	Freshly prepared juice	Boiled for 3 minutes
Orange	142.07 mg/ 100 g	123.63 mg/ 100 g
Guava	307.38 mg/ 100 g	282.79 mg/ 100 g
Lemon	103.14 mg/ 100 g	98.361 mg/ 100 g

Storage at ambient temperature reduces the ascorbic acid concentration:

To retain the concentration of ascorbic acid in fruits, it is important to store fruits at an optimum temperature. To study the effect of storage conditions on ascorbic acid concentration, redox titration was carried out on two experimental groups. The first group consisted of freshly prepared juice of orange, guava, and lemon. The second group consisted of fruit juice samples, which were stored at 28°C for 3 days. It was observed that there was a reduction in ascorbic acid concentration in 3-day stored samples as compared to freshly prepared fruit juices (Figure 7 and Table 3). The concentration of ascorbic acid was estimated to be 112.7 mg/ 100 g for the orange juice sample, 282.79 mg/ 100 g for the guava juice sample, and 96.311 mg/ 100 g for the lemon juice sample. Exposure to the atmosphere in ambient conditions reflects a degradation of ascorbic acid, and the extent of degradation was 20.7% for orange, 8% for guava, and 6.6% for lemon.

DOI: 10.36838/v7i10.31

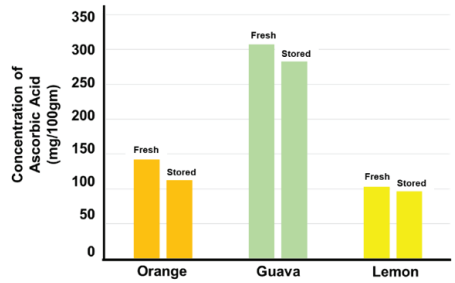


Figure 7: Storage at ambient temperature degrades ascorbic acid concentration in fruit juice samples. The bar graph represents the concentration of ascorbic acid (mg) per 100 g of fruit sample in freshly prepared and stored juices at ambient temperature, 28 °C, for 3 days of orange, guava, and lemon.

Table 4: Concentration of ascorbic acid in 100 g of freshly prepared fruit samples and 100 g of fruit samples stored at an ambient temperature of 28 °C for 3 days.

Fruit Sample	Concentration of ascorbic acid (mg) per 100g of sample	
	Freshly prepared juice	Stored for 3 days at 28°C
Orange	142.07 mg/ 100 g	112.7 mg/ 100 g
Guava	307.38 mg/ 100 g	282.79 mg/ 100 g
Lemon	103.14 mg/ 100 g	96.311 mg/ 100 g

Discussion

Vitamin C plays a crucial role in forming collagen, supporting immune functions, and cyto-protection. Our body cannot produce Vitamin C; thus, we need to be dependent on different dietary sources of Vitamin C, like citrus fruits, vegetables, etc. Vitamin C is an extremely sensitive Vitamin, and thus maintaining optimum temperature and storage conditions is crucial to retain the concentration of this Vitamin. The present study aims to estimate the concentration of Vitamin C in different fruit samples and to study the effect of temperature and storage conditions on the concentration of Vitamin C.

The study shows that the concentration of ascorbic acid (Vitamin C) varies across different fruits. The study showed that guava has the highest concentration of Vitamin C (307.38 mg/ 100 g), followed by orange (142.07 mg/100 g) and lemon (103.14 mg/100 g). In prior studies, the Vitamin C content in orange was estimated to be (58.304 mg/ 100 g) by Najwa *et al.*,[6] (42.7 mg/100 g) by Tee *et al.*,[7] (56.020 mg/100 g) by Cioroi *et al.*,[8] and (56.4 mg/100 g) by Bungau *et al.*[9] The concentration of Vitamin C in orange is lower by 59%-70% across these aforementioned studies compared to the observed concentration in these experiments. Lack of standardization of fruit sourcing (species, extent of ripening, storage conditions) as well as different testing methodologies (High Performance Liquid Chromatography[6], Microfluorometry[7], dye-based titration[6]) are reasons for this variation compared to earlier studies. For guava, the Vitamin C content was estimated to be (206-334 mg/100 g) in prior studies by Ghani *et al.*,[10] (330.77 mg/100 g) by Alok *et al.*,[11] and (181.6 mg/100 g) by Simran *et al.*[3] in prior studies. The observed concentration of Vitamin C in guava is lower by 12% compared to the estimates by Ghani *et al.*[10] and higher by 7% compared to the estimates by Alok *et al.*[11].For lemon, the Vitamin C content was estimated to be (43.956 mg/100 g) by Najwa *et al.*,[5] (46.8 mg/ 100 g) by Tee *et al.*,[7] (51.78 mg/100 g) by Cioroi *et al.*,[8] and (49 mg/100 g) by Bungau *et al.*[10] in earlier studies. The observed concentration of Vitamin C in lemon is lower by 57% compared to the estimates by Najwa *et al.*,[5] lower by 55% compared to the estimates by Tee *et al.*,[7] lower by 50% compared to estimates by Cioroi *et al.*,[8] and lower by 52% compared to estimates by Bungau *et al.*[10] These differences are due to the lack of standardization of the fruit sources, stage of ripening, and storage conditions of the different samples. The testing methods used across these studies are also varied (High Performance Liquid Chromatography,[6] Microfluorometry,[7] dye-based titration.[6] However, guava has been shown to have the highest concentration of ascorbic acid across these three fruits, and this is consistent with prior studies.

All the citrus fruits show a degradation of ascorbic acid even when subjected to short durations of boiling temperatures. Prior studies also demonstrated similar results where boiling fruit juices resulted in the reduction of Vitamin C content. In an earlier study, a 10% reduction of ascorbic content was shown when orange samples were boiled/ exposed to heat.[12] Similarly, a degradation of 18% was reported for guava samples at boiling temperatures.[13] In the case of lemon, earlier studies by Njoku *et al.*[12] have shown a reduction of 18%, while another study showed a reduction of 13% when exposed to a boiling temperature.[14] The extent of degradation is directionally in line with other studies; however, other studies did pursue the extent of degradation for much larger time durations. Ascorbic acid is sensitive to temperature, both while cooking, food processing, as well as transport and storage conditions, and this is reflected in the varied levels of ascorbic acid when subjected to higher temperatures.

Storage conditions in the open also lead to degradation of ascorbic acid, resulting in reduced concentrations across fruits. In a prior study, reduction of ascorbic acid was shown to the extent of 16.27% when orange samples were stored.[15] For guava, this was reported to be 28% in a prior study.[16] For lemon, there was a reduction of 10% reported in an earlier study.[17] It is to be noted that the duration of storage was varied in the prior studies, but all studies were consistent in the reduction of ascorbic acid over time. Thus, the present study indicates that the concentration of Vitamin C varies in different samples, and high temperature and adverse storage conditions reduce the concentration of this Vitamin.

The limitations of this study are the lack of experimental replicates and statistical analysis. This study does not have enough experimental data points for robust statistical analysis. It is recommended that, for more robust data and analysis, experiments should be performed in triplicate and should include basic statistical analysis (means, standard deviations, statistical significance).

Conclusion

Vitamin C (ascorbic acid) is an important micronutrient that plays a crucial role in maintaining human health. Dietary guidelines recommend the consumption of Vitamin C-rich fruits to meet daily dietary requirements. The present findings demonstrate that the concentration of Vitamin C is influenced by external factors such as the source, storage, and transport conditions of the fruit. Boiling for 3 minutes and incubation

of fresh juice for 3 days at an ambient temperature of 28 °C significantly reduced Vitamin C concentration. Optimum temperature and storage conditions are critical for preserving Vitamin C content in fruits since this vitamin is highly sensitive to heat, light, and oxygen. Thus, optimum temperature and storage conditions should be maintained to reduce degradation of Vitamin C, thereby reducing the nutritional value of the fruit.

The practical implications of this study can broadly be summarized into three broad areas: choice of fruits for Vitamin C supplementation, cooking and food processing implications, and storage implications. Firstly, from a nutritional supplementation perspective, guava, followed by orange, and subsequently lemon, provide the greatest amount of Vitamin C supplementation based on the results of the study. These fruits can therefore be the choice of Vitamin C supplementation based on patient/ consumer needs. Secondly, cooking and exposure of these fruits to boiling temperatures results in degradation and lesser availability of Vitamin C. High temperature food processing should be avoided on these fruits, to maximize nutritional availability of Vitamin C. Thirdly, the storage of these fruits for extended periods at 28 °C or higher results in degradation of Vitamin C. These fruits should ideally be kept in cold storage to maximize the concentration of Vitamin C in these fruits, and should be consumed immediately after being removed from cold storage.

Acknowledgments

I would like to acknowledge the support of my family, who supported me through this project.

References

1. Doseděl, M., *et al.* (2021). Vitamin C—Sources, Physiological Role, Kinetics, Deficiency, Use, Toxicity, and Determination. *Nutrients*, 13(2), 615. https://doi.org/10.3390/nu13020615
2. Lv, X., *et al.* (2015). Citrus fruits as a treasure trove of active natural metabolites that potentially provide benefits for human health. Chemistry Central journal, 9, 68. https://doi.org/10.1186/s13065-015-0145-9.
3. Dani, S. (2023). A study on physical and chemical characteristics of guava (*Psidium guajava L.*), Taiwan pink variety. Pharma Innovation, 2023, 12(5), 2023-2025. https://doi.org/10.22271/tpi.2023.v12.i5y.20175
4. Aliansa, T., *et al.* (2023). The determination of Vitamin C in guava (Myrtaceae species) using spectrophotometric approach. Asian Journal of Analytical Chemistry, 1(1), 6-11. https://doi.org/10.53866/ajac.v1i1.268
5. Rameshwari, K., Sudha, P., A., Raksitha, T., SANGEETHA, K., & J. Vijayashanthi. (2022). A comparative study on the ascorbic acid content of several fruits, International Journal of Food and Nutritional Sciences, 2022, 1(7), 1224-1232
6. R. Fatin Najwa, & A. Azrina. (2017). Comparison of Vitamin C content in Citrus fruits by Titration and High-performance Liquid Chromatography (HPLC) methods. International Food Research Journal, 24(2), 726-733
7. Tee, E. S., Young, S. I., Ho, S. K., & Shahid, S. M. (1988). Determination of Vitamin C in fresh fruits and vegetables using Dye-titration and Microfluorometric Methods. Pertanika, 11(1), 39-44
8. Cioroi, M. (2007). Study on L-ascorbic acid contents from exotic fruits. *Cercetari Agronomice Moldova, 1(129), 23-27
9. S. Bungău, *et al.* (2011). Studies on citrus species fruits ascorbic acid content using kinetic, spectrophotometric and iodometric methods. Analele Universităţii Din Oradea, Fascicula: Protecţia Mediului, 16(16), 212-217.
10. Abdul Ghani, T. H., Hussain, M., Ikram, M., Imran, M., Farooq, M., & Muhammad, N. (2016). Study on chemical changes of different guava varieties during different ripening stages. International Journal of Basic and Applied Chemical Sciences, 6(1), 38-40
11. Alok, R. N., Chatterjee, D. D., Roy, T., Hossain, A. M. Z., & Azizul-Haque, M. D. (2011). Study on chemical changes of different guava varieties during different ripening stages. Bangladesh Research Publications Journal, 6(2), 217-224
12. Njoku, P. C., *et al.* (2011). Temperature effects on Vitamin C content in citrus fruits. Pakistan Journal of Nutrition, 10(12), 1168-1169. https://doi.org/10.3923/pjn.2011.1168.1169
13. N, M., & JYOTHI, B. (2016). Colorimetric determination of Vitamin C in fresh and dilute fruit juices and effect of thermal exposure on concentration at various stages. International Journal of Pharma and Bio Sciences, 7(4). https://doi.org/10.22376/ijpbs.2016.7.4.p197-211
14. Sehgal, A., Jhanwar, B., & Gilhotra, U. (2015). Active content variation in citrus lemon: Age, temperature, pH, and air. Frontiers in Food & Nutrition Research, 1(1), 1-10
15. CARPIUC, N., LEAHU, A., CURALEŢ, E., & DAMIAN, C. (2011). EFFECT OF STORAGE ON ASCORBIC ACID CONTENT OF SOME FRUIT JUICES. Food and Environment Safety Journal, 10(4), 95-99
16. Ali Asad Yousaf, *et al.* (2024). Storage stability assessment of guava fruit (*Psidium guajava L.*) cv. "Gola" in response to different packaging materials. Sustainable Food Technology, 2, 210-221. https://doi.org/10.1039/d3fb00113j
17. Sinha, A. (2014). Studies on ascorbic acid (Vitamin C) content in different citrus fruits and its degradation during storage. Science and Culture, 80 (9-10), 265-268

Author

Dhruv Prashant is a Grade 11 student researcher with a deep interest in biochemistry and molecular biology. He is committed to serving the underprivileged, leveraging the transformative power of science and medicine. He seeks to pursue a career in medicine that can uplift society by improving human health.

DOI: 10.36838/v7i10.31

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Investigating the Combinatorial Effects of Nanoparticles and Antioxidant Compounds on *Drosophila Melanogaster*

Andrew G. Mozinski

Crimson Global Academy, 18 Stanley Street, Level 3, Parnell, Auckland, 1010, New Zealand; drewmozinski@gmail.com

ABSTRACT: Nanoparticles have been reported to potentially cause a detrimental effect on fertility and physical fitness in humans. This is a concern due to an increasing exposure of nanoparticles in our daily environment. Previous studies have also shown that antioxidant compounds increased longevity and fertility in model organisms. In this study, we used a *Drosophila melanogaster* model to examine the combinatorial effects of nanoparticles and antioxidants. First, using a climbing motility assay and a larvae count fertility assay, we looked at three commonly found nanoparticles: gold, silver, and titanium, and established a working concentration where they exerted a negative influence on the physical fitness and fertility of *Drosophila* adult flies. We then determined the beneficial effect of antioxidants melatonin, N-acetylcysteine (NAC), and quercetin. Subsequently, we examined the combined effects of nanoparticles and antioxidants on *Drosophila* physical fitness and fertility. Furthermore, we examined the transgenerational effect of these chemical combinations on the first generation of progeny flies. In our studies, we found that, to varying degrees, each nanoparticle and antioxidant has a detrimental or beneficial effect, respectively, on both mobility and fertility of *Drosophila melanogaster* transgenerationally, with silver nanoparticles being the most detrimental, and melatonin being the most beneficial.

KEYWORDS: Biochemistry, Medical Biochemistry, Nanoparticles, Antioxidants, *Drosophila melanogaster*.

Introduction

Nanoparticles (NPs) have biological impacts that raise sizable concerns due to their increasing prevalence in industrial, medical, and consumer applications. While several nanoparticles are biocompatible and have revolutionized medical and industrial fields, metal-based nanoparticles have displayed potentially hazardous effects on biological systems. These particles, as a result of their minuscule size, have the potential to disrupt crucial physiological processes via direct interaction with cellular structures.[1-3] Research has shown that metal-based nanoparticles, such as gold (Au), silver (Ag), and titanium dioxide (TiO2), can exert detrimental effects on the fertility, development, and overall health of organisms such as *Drosophila melanogaster*. [4-7]

Drosophila melanogaster was the ideal model organism for this project for various reasons. Its short life cycle, fecundity, and genomic structure have made it desirable for a large portion of toxicological studies.[8] However, its most attractive trait for this project was its commercial availability to the general public. *Drosophila melanogaster* is useful for observing the toxicological impacts of NPs and has demonstrated severe developmental delays in response to exposure to TiO2 NPs.[9,10] Previous research has also shown that larval stages of *Drosophila melanogaster* are particularly susceptible to nanoparticle-induced genotoxicity.[11] This genotoxicity can result in aberrant phenotypes, as well as increased susceptibility to DNA damage and behavioral changes that could alter reproductive efficacy over generations. These findings are concerning, especially concerning the safety considerations of these materials in consumer products.

Numerous antioxidants have been the subject of the limelight in recent years. Commercially available compounds such as N-acetylcysteine (NAC), melatonin, and quercetin show promise in oxidative stress mitigation.[12, 13] NAC acts as a glutathione precursor and is applicable in detoxification.[14,15] Melatonin and quercetin, in addition to their antioxidant properties, have been found to ameliorate the neurotoxic effect induced by TiO2 NPs exposure, potentially allowing them to have a similar impact on the neurotoxic effect of AgNPs.[16-19]

This study aims to evaluate and compare the effects of Au, Ag, and TiO2 nanoparticles on the physical fitness and fertility of *Drosophila melanogaster*, while simultaneously measuring the extent to which antioxidants, NAC, melatonin, and quercetin, can mitigate the adverse effects. We hypothesize that Beneficial compounds such as NAC, quercetin, and melatonin can reverse the detrimental effects from nanoparticles such as silver, gold, and titanium dioxide, on fertility and physical capabilities of *Drosophila melanogaster*. Using established mobility and fertility assays, the study aims to contribute to a deeper understanding of nanoparticle toxicity and possible counteractants.

Methods

Drosophila melanogaster maintenance:

Wingless Drosophila melanogaster specimens were obtained from a commercial supplier, Josh's Frogs, and maintained under strictly controlled environmental conditions (Figure 1). The fruit flies are maintained on a culture medium that was obtained from the same manufacturer. Josh's Frogs Culture media is primarily powdered-potato-based, while using vitamins and antifungal ingredients. The powdered culture media

DOI: 10.36838/v7i10.36

were reconstituted with water according to the manufacturer's instructions. The flies were housed at a steady temperature of 25°C with a photoperiod consisting of 16 hours of light and 8 hours of dark. A regulated thermal environment was established inside a 20-gallon glass reptile terrarium equipped with a standard heating apparatus. Humidity of the thermal environment was maintained at 65%. Temperature and humidity were monitored using a digital thermometer and a smart hygrometer used in conjunction.

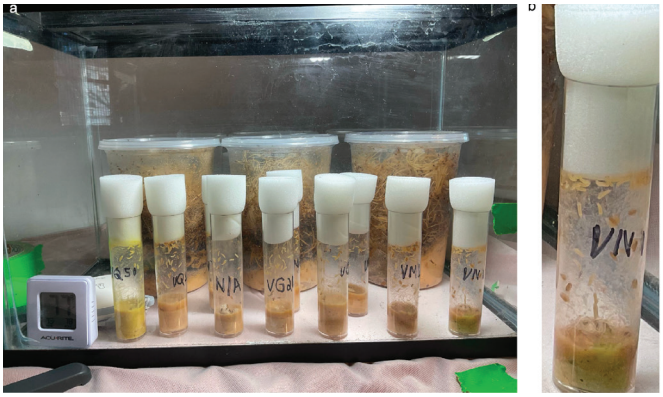


Figure 1: Images of the experimental set up.(a) A controlled tank environment at 25 deg Celsius containing culture vials labelled by food conditions. (b) A culture vial containing food culture and *Drosophila melanogaster*.

Reconstitution of chemicals:

To assess the impact of the various chemicals, melatonin (Sierra Life Sciences), N-Acetyl L-Cysteine (NAC) (Bulk-Supplements.com), quercetin dihydrate (Micro Ingredients), gold (Au) nanoparticles (NutriNoche), silver (Ag) nanoparticles (XFNANO), and titanium dioxide (TiO2) nanoparticles (Alexes) were obtained from reputable vendors in powdered or liquid form. From a literature search, three different concentrations were tested for each chemical.[9,21-25] The chemicals are mixed thoroughly into the culture media reconstituted as described above. Three replicates were performed for each chemical condition using identical 95mm polystyrene culture vials. Chemicals were dissolved in water as a stock solution. Serial dilutions were then used to create solutions of applicable concentrations.

Selection of optimal chemical concentrations:

The prepared solutions were thoroughly incorporated into the culture media within the designated culture vials, ensuring uniform exposure for all experimental groups. Experimental groups were composed of ten flies per culture vial, with a 5:5 male-to-female ratio. The flies were placed into 95mm polystyrene vials containing a culture medium supplemented with precise concentrations of experimental compounds. These vials were labeled according to their experimental conditions and maintained within the controlled environment. Following the initial exposure period, the optimal concentration was determined for each chemical based on maximum observable effects on the flies while minimizing lethality (Table 1). Only the optimal conditions (highest concentration with no observed fly fatalities) of each chemical were used for further studies. Vials that contained suboptimal concentrations were not analyzed further. This process ensured that all experiments utilized the most effective dosages for measurement (Table 2)

Table 1: List of chemicals used and the concentrations tested. The highest concentration with no fatalities observed were used in subsequent experiments.

Chemical used	Concentration	Fatalities observed	Best?
NONE (Food only)	NONE (Food only)	No	**Yes**
NAC	0.1 mg/mL	No	**Yes**
NAC	1 mg/mL	Yes	No
NAC	5 mg/mL	Yes	No
Quercetin	0.1 mg/mL	No	**Yes**
Quercetin	0.5 mg/mL	Yes	No
Quercetin	5 mg/mL	Yes	No
Melatonin	0.01 mg/mL	No	No
Melatonin	0.1 mg/mL	No	**Yes**
Melatonin	0.5 mg/mL	Yes	No
Titanium	0.0014 mg/mL	No	**Yes**
Titanium	0.014 mg/mL	Yes	No
Titanium	0.14 mg/mL	Yes	No
Gold	1 ppm	No	No
Gold	20.5 ppm	No	**Yes**
Gold	30 ppm	Yes	No
Silver	0.1 mg/mL	No	**Yes**
Silver	1 mg/mL	No	No
Silver	10 mg/mL	Yes	No

Table 2: Combinations of Antioxcidants and nanoparticles used.

Condition	Reversal Compound	Nanoparticle
ALPHA	FOOD ONLY	FOOD ONLY
A	NAC	Titanium
B	NAC	Gold
C	NAC	Silver
D	Quercetin	Titanium
E	Quercetin	Gold
F	Quercetin	Silver
G	Melatonin	Titanium
H	Melatonin	Gold
I	Melatonin	Silver

Experimental Groups:

There are four main experimental groups used in this experiment, each of which is staggered apart by a week (Figure 2). P1 refers to the first experimental group for the parent generation and contains all of the individual culture vials that test how each chemical alone affects *Drosophila melanogaster* (Figure 2, Table 1). Three replicates were performed for each chemical condition. P2 refers to the second experimental group of the parent generation, which is subjected to a different combination of chemicals (Figure 2, Table 2). The combined conditions examine whether each nanoparticle's effect on *Drosophila melanogaster* can be mitigated by each antioxidant reversal compound. Due to the large number of permutations, only two replicates of each vial are used in P2. F1 refers to the first experimental group of the filial generation. F1 consists of the offspring of P1 that matured in an environment with only food and no added chemicals. The goal is to measure the potential transgenerational effects that each chemical has on the progeny after the parents were exposed to chemicals. The final group is F2, the second group in the filial generation. F2

is the offspring of P2 that matured in a food-only medium with no chemicals. Thus, this condition examines whether the transgenerational effect of each nanoparticle on *Drosophila melanogaster* can be mitigated by the antioxidant reversal compounds. All experimental groups contain at least one control vial that only contains regular food. The conditions with different combinations of chemicals are labeled A-I in Table 2.

Physical mobility assay:

The locomotor activity of the flies was assessed daily for five days. Five randomly selected files from each condition were placed into a fresh empty vial, where they were allowed a brief acclimatization period of 5 minutes. Following acclimatization, the vial was gently tapped so that the flies were at the bottom of the vial. The number of flies that successfully climbed to three-quarters of the vial's height within 1 minute was recorded. This procedure is repeated for replicates of each condition daily to monitor the changes over 5 days.

Fertility Assay:

Fertility was assessed after the physical mobility assay. On the sixth day, all adult flies are removed from each experimental group, and the larval progeny are counted to assess reproductive output. For the parent generation, the adults are placed into clean culture vials containing normal culture medium and given 4 days for reproduction before being removed and disposed of. The new larvae in the food-only vials are given 8 days to develop into adults before being transferred, in groups of ten with a 5:5 male-to-female ratio, into new culture vials. These new cultures are the first filial generation. They are then assessed under the same 6-day timeline as the previous generation to assess fertility. The remaining adult and larva vials are disposed of. All culture vials containing flies, after their usage, were frozen overnight, wrapped in disposable paper towels, and discarded via standard disposal channels.

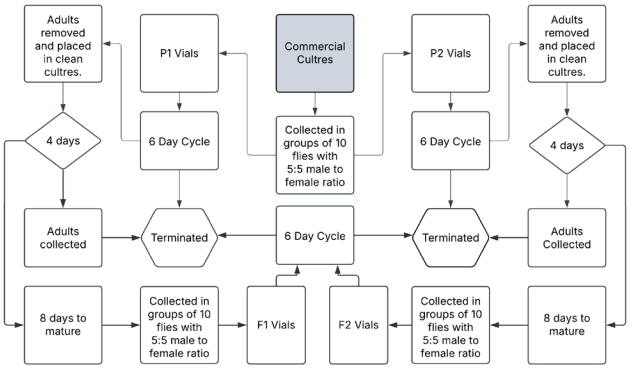


Figure 2: Flowchart illustrating experimental procedures and naming of different groups of files.

Results and Discussion

Optimization of Chemical Concentrations to be used:

The optimal concentrations of each chemical used in the experiment were determined based on the largest impact without fatalities. Table 1 shows all chemical concentrations tested in the initial exposure. Even for the beneficial antioxidant compounds melatonin, NAC, and quercetin, all chemicals resulted in fatalities when administered in doses of high concentrations. This initial exposure was done using the same flies that would later be assayed. Suboptimal concentrations and their duplicates were subsequently disposed of prior to the start of the experiment. The optimal chemical concentrations utilized for the subsequent experiments are as follows: melatonin 0.1mg/mL, N-Acetyl L-Cysteine (NAC) 0.1mg/mL, quercetin dihydrate 0.1mg/mL, gold (Au) nanoparticles 20.5ppm (or 0.025mg/mL), silver (Ag) nanoparticles 1mg/mL, and titanium dioxide (TiO2) nanoparticles 0.00014mg/mL (Table 1).

Group P1 – Mobility Assay:

In the control group of the first parent generation P1, 4 out of 5 flies climbed to ¾ of the tube in the mobility assay, which is in line with the standard geotaxis of *Drosophila melanogaster*. Flies kept in food containing antioxidants showed a higher proportion of flies climbing than the control group, with melatonin being the highest by far in the initial days, followed by NAC and quercetin. Flies in all 3 antioxidant groups showed better mobility performance than the control group.

For the nanoparticles, flies grown in AuNPs conditions performed the worst out of all conditions, with as low as an average of 2 displaying geotaxis during the later days of their testing cycle. TiO2 had similar results to AuNPs; however, it did not decline as rapidly in its later days. Silver had the least effect on *Drosophila melanogaster* as files in the AgNPS showed the highest mobility out of all 3 NPS conditions, but still lower than the control group (Figure 3). This was surprising as silver's neurotoxic effect was reported to have a significant detrimental effect on the flies, especially at the concentration it was administered.

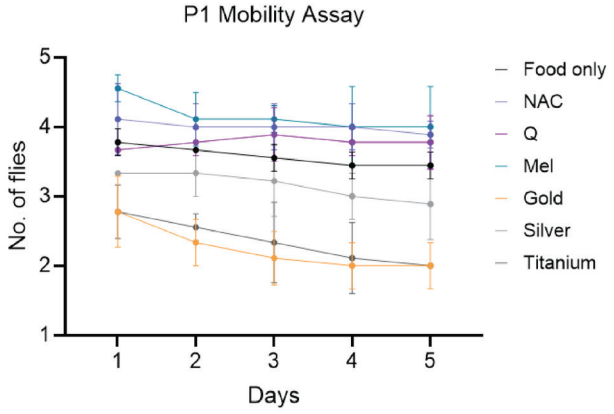


Figure 3: Mobility assy of parent flies exposed to individual compounds. Files exposed to nanoparticles exhibit significant decreased mobility from control (food only) except silver. For antioxidants, no significant changes were observed except in the condition with melatonin.

DOI: 10.36838/v7i10.36

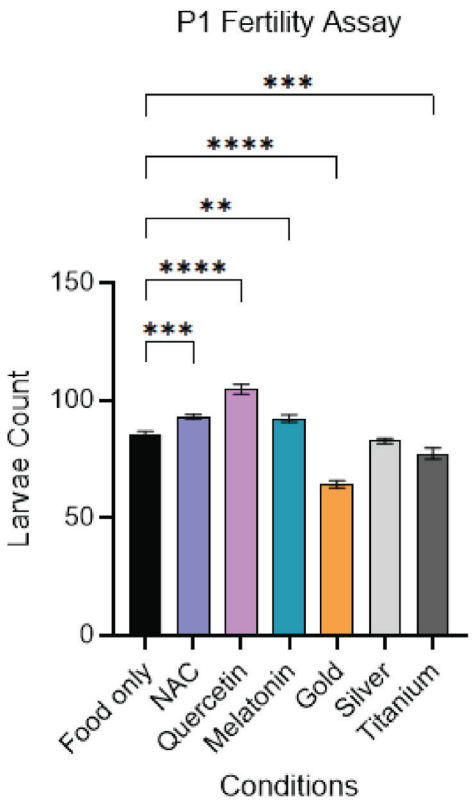


Figure 4: Fertility assay counting larvae from parent flies exposed to individual compounds. Flies exposed to nanoparticles exhibit significant decrease from control, while flies exposed to antioxcidants exhibit a significant increase. Statistical analysis was performed using one-way ANOVA followed by post-hoc Turkey's test with ** p 0.01, ***p 0.001 and ****p 0.0001.

Group P1 – Fertility Assay:

In the fertility assay, flies in the antioxidant groups produced more larvae than the control group, with the quercetin condition producing the highest number of larvae, followed by melatonin and NAC. On the other hand, files in NPS groups predictably produced fewer larvae, with gold having the largest detrimental effect of reducing the average larvae count to 64.3 (Figure 4).

Group P2 – Mobility Assay:

The second experiment, done using the parent generation, is named P2. In this experiment, flies are exposed to different combinations of nanoparticles (Table 2). Flies in the different combinations performed similarly in the mobility assay. When interpreted in conjunction with the P1 data, we observed that despite being the most detrimental nanoparticle when exposed individually, the effect of gold was greatly mitigated by the antioxidants. The antioxidant that was most effective in counteracting the oxidative stress caused by the AuNPs was melatonin, restoring flies' locomotion to even above the control group during the initial days of the experiment. Following melatonin, the NAC, and Quercetin groups also significantly decreased the mobility defect from oxidative stress that AuNPs caused (Figures 5, 6).

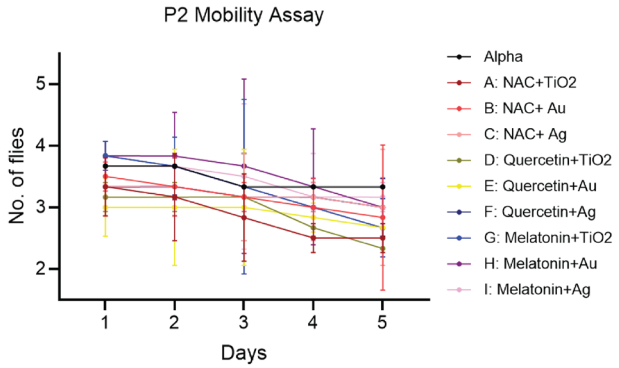


Figure 5: Mobility of parent flies exposed to combinations of compounds. All combinations demonstrate a non-significant difference, suggesting a revearsal effect from the antioxidants on the detrimental effect from exposure to nanoparticles.

TiO2 had similar results with some notable discrepancies: Melatonin rescued some locomotion defects, but overall mobility was still lower than the control group. NAC and quercetin were similar in terms of mitigation, with NAC still being more effective. We also observed that while the effect of TiO2 is mitigated by the antioxidants, the mitigation effect was less than what was observed in AuNPS (Figure 5,7).

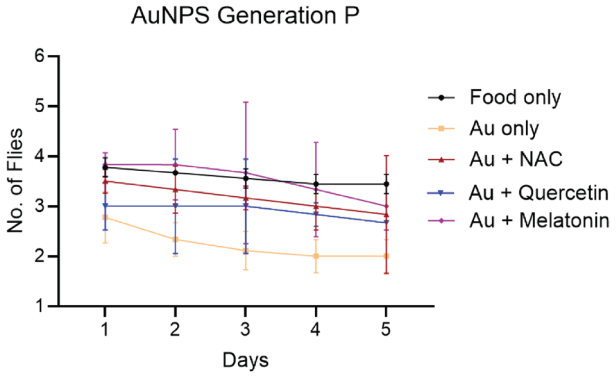


Figure 6: Mobility assay of parent flies exposed to AuNPs in combination with different antioxidants. The AuNPs-only group showd the lowest mobility, while every other combination showed a non-significant difference from the control. This suggests a reversal effect from the antioxidants in mediating the effect of AuNPs.

The combination of AgNPs and antioxidant exposure yielded the most interesting results. While there was no noticeable mitigation from either quercetin or NAC, the addition of melatonin improved mobility to almost the same extent as the TiO2 group, mitigated by melatonin (Figures 7, 8). This suggests that melatonin might be able to counteract the effects of Silver due to its protective properties against neurotoxicity in addition to its antioxidant properties. We speculate that silver has a neurotoxic effect on *Drosophila melanogaster*, causing a decrease in fitness. This would also explain why quercetin consistently performed marginally better than NAC despite causing worse mobility in the individual conditions group, as quercetin has also been linked to ameliorating neurotoxic effects. This would also give an insight into the marginally worse performance of the antioxidants in counteracting TiO2, which

has a verified neurotoxic effect as well, due to triggering neuroinflammation.

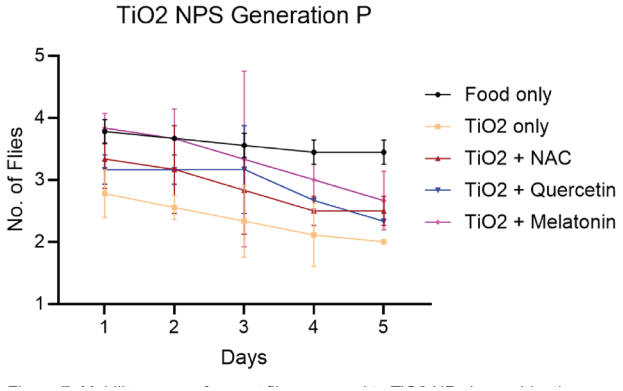


Figure 7: Mobility assay of parent flies exposed to TiO2 NPs in combination with different antioxidants. The TiO2 NPs-only group showed the lowest mobility, while every other combination showed a non-significant difference from the control. This suggests a reversal effect from the antioxidants in mediating the effect of TiO2 NPs.

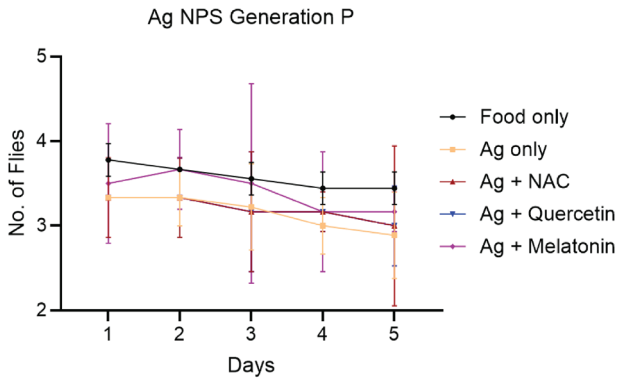


Figure 8: Mobility assay of parent flies exposed to Ag NPs in combination with different antioxidants. The Ag NPs-only and the Ag + NAC combination group showed the lowest mobility, while every other group showing a non-significant difference from the control. This suggests a reversal effect from the antioxidants, while also demonstrating silver's resistance.

Group P2 – Fertility Assay:

In the fertility analysis of P2, we observed that all Groups A-H showed lower larval count than the control. Group I (AgNPS + Melatonin) was the only group with a larval count of 83.5, close to the control group larval count of 84. Compared to other oxidants (Groups C and F), melatonin is potentially more effective in counteracting the negative effects of AgNPs. Other groups performed similarly, with larval count lower than the control group. Groups treated with melatonin (Groups G, H, I) generally showed a higher larval count (Figure 9).

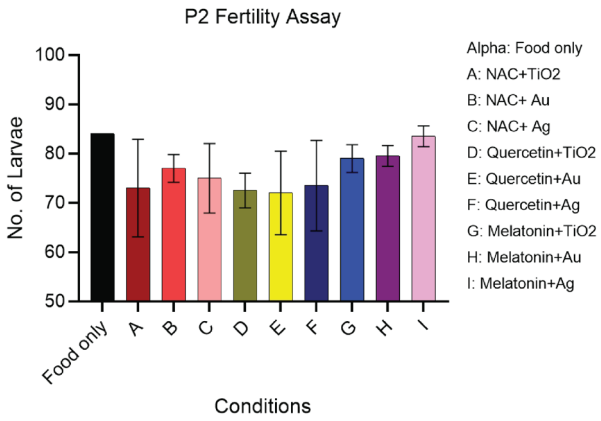


Figure 9: Mobility assay of parent flies exposed to different chemical combinations. All combinations demonstrate a non-significant difference compared to the control, indicating a reversal effect from the antioxidants. Statistical analysis was performed using one-way ANOVA followed by post-hoc Tukey's test.

Group F1 – Mobility Assay:

In the F1 group, we examined the first generation of offspring for mobility effects after the parent generations had been exposed to the individual compound. Offspring flies from conditions containing AuNPS, AgNPS, and TiO2 continued to show reduced mobility. Offspring from the AgNPS condition have the lowest mobility. On the other hand, offspring flies from Quercetin, melatonin, and NAC showed mobility similar to the control group (Figure 10).

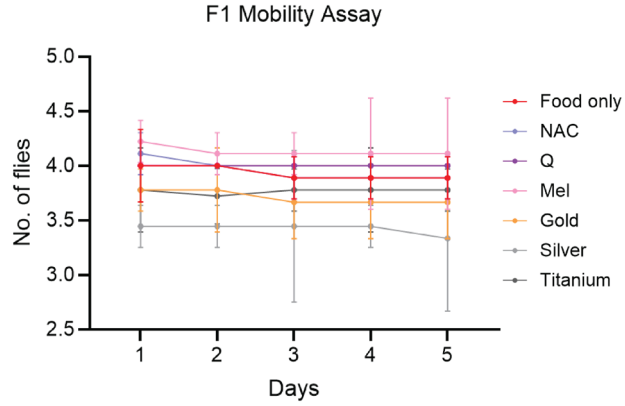


Figure 10: Mobility assay of spring flies from parent flies exposed to individual compounds. Offspring of parents exposed to melatonin, gold, and titanium showed significant differences from control. Silver also demonstrated a significant decrease from the control.

Group F1 – Fertility Assay:

In the F1 group, offspring flies from the quercetin and melatonin conditions showed a noticeable increase in larval count compared to the control group. On the other hand, NAC and TiO2 conditions have larvae count similar to the control group. The largest decrease in larval counts was observed in the offspring from the AuNPS and AgNPS conditions. This suggests to us that the detrimental impact of AgNPs and AuNPs is exerted transgenerationally on the fertility capability of the offspring (Figure 11). Together with the mobility assay data, AgNPs and AuNPs seem to exert detrimental transgenerational effects on both the fertility and mobility of *Drosophila*

DOI: 10.36838/v7i10.36

melanogaster (Figures 10, 11). Melatonin and quercetin, on the other hand, also have a transgenerational impact, with melatonin affecting fertility and mobility, and quercetin impacting mainly fertility (Figures 10, 11).

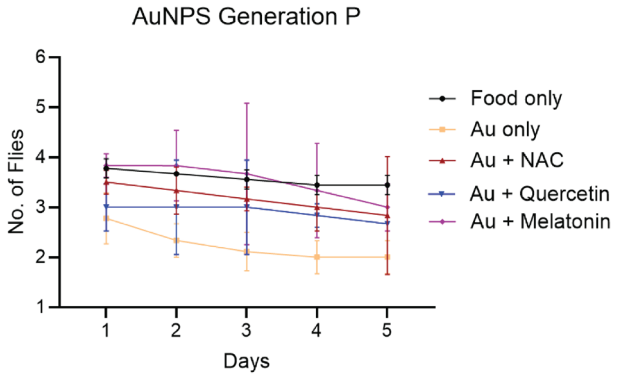


Figure 11: Fertility assay of offspring flies from parent flies exposed to individual compounds. Flies from parents exposed to nanoparticles exhibit a significant decrease from control, except for TiO2 NPs. Flies from parents exposed to antioxidants exhibited a significant increase from the control, except for NAC. Statistical anlaysis was performed using one way ANOVA followed by post-hoc Tukey's test with ***p<0.0001.

Group F2 – Mobility Assay:

In the F2 group, we examined the first generation of offspring for mobility effects after the parent generations had been exposed to combinations of different compounds (Figure 12). The F2 group yielded interesting and exciting results. While the Au nanoparticles performed lower in the F1 group, all groups treated with antioxidants performed similarly to the control group in mobility assays (Figure 13). For both AuNPs and TiO2 nanoparticles conditions, offspring flies exhibit a higher performance in mobility with all three antioxidants (Figures 14, 15). Despite not demonstrating a clear transgenerational impact on F2 flies, TiO2 still scored lower in all categories when compared to the antioxidant-treated vials in F2. AgNPs were the only nanoparticles that seemed to display persistent effects despite being treated with antioxidants. Both the quercetin and the NAC-treated silver groups showed no discernible improvements in mobility. However, the melatonin-treated groups did score very close to the control group in mobility assessments. These results are significant because they suggest a clear reversal of the transgenerational damage caused by every nanoparticle. This data suggests that the antioxidants tested can reverse the transgenerational damage caused by both titanium and gold nanoparticles, while melatonin appears to have the unique ability to counteract the effects of silver nanoparticles, likely due to its research link to counteracting neurotoxic effects (Figures 12, 13, 14, 15).

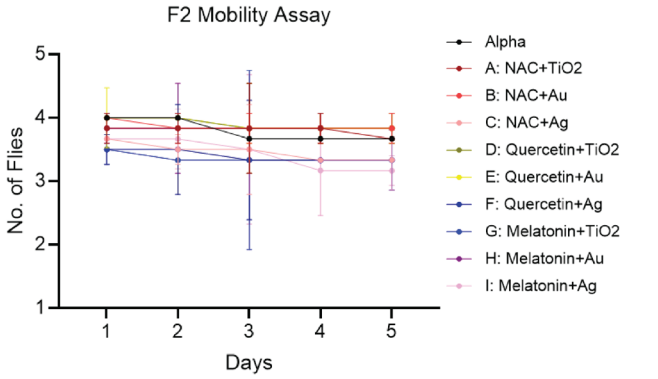


Figure 12: Mobility assay of offspring flies from parent flies exposed to combinations of chemicals. All combinations except C, F, and G demonstrate a non-significant difference compared to control, suggesting a reversal effect from the antioxidants. Conditions C, F, and G showed a significant decrease from the control. Statistical analysis was performed using one-way ANOVA followed by post-hoc Tukey's test.

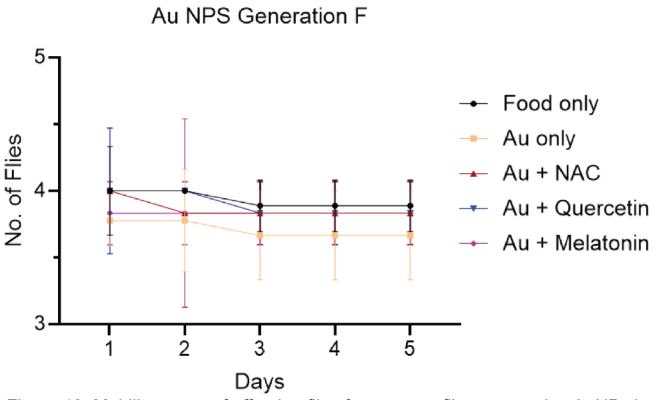


Figure 13: Mobility assay of offspring flies from parent flies exposed to AuNPs in combination with different antioxidants. The AuNPs-only group showed the lowest mobility, with other combinations showed a non-significant difference from the control. This suggests a reversal effect from the antioxidants.

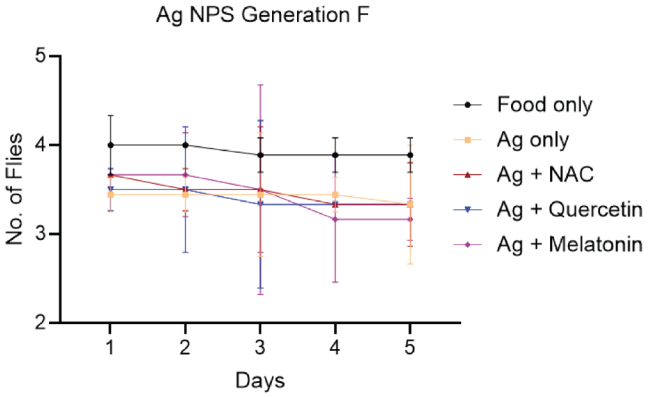


Figure 14: Mobility assay of offspring flies from flies exposed to AgNPs in combination with different antioxidants. The AgNPs-only group showed the lowest mobility. Other combinations except for Melatonin showed a significant decrease from the control.

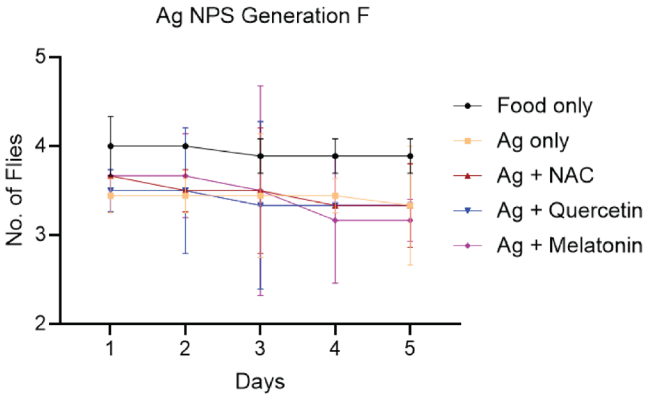


Figure 15: Mobility assay of offspring flies from parent flies exposed to TiO2 NPs in combination with different antioxidants. No combinations showed any significant difference from the control.

Group F2 – Fertility Assay:

In the F2 group, offspring flies from the TiO2 groups, when combined with NAC and Melatonin (Figure 16, Conditions A and G), showed a recovery of fertility, while the combination with Quercetin did not (Figure 16, Condition D). The offspring flies from the AuNPs group in combination with NAC and quercetin (Figure 16, Conditions B and E) exhibited recovered fertility compared to the control group. AuNPs in combination with melatonin did not show a recovery in fertility rate (Figure 16, Condition H). Lastly, offspring flies from AgNPs, when in combination with quercetin and NAC (Figure 16, Conditions C, F), did not show any recovery of fertility. However, the exposure of melatonin in combination with AgNPs results in a fertility rate close to that of the control group.

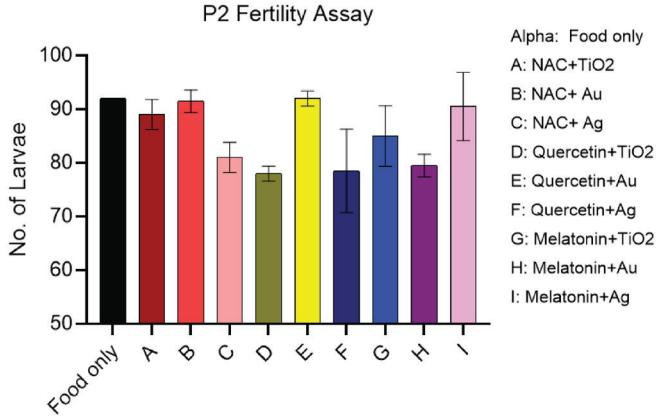


Figure 16: Fertility assay of offspring flieds from parent flies exposed to different chemical combinations. All combinations demonstrate a nonsignificant difference compared to the control, suggesting a reversal effect from the antioxidants.

Results and Data Analysis

In our studies, we observed that each nanoparticle has exerted a detrimental effect on both the mobility and fertility of Drosophila melanogaster. Residual effect was still observed in the next generation on mobility, and to a lesser extent, fertility. Interestingly, the reverse was observed on antioxidant compounds. Antioxidants improved both mobility and fertility in the parent flies, and quercetin and melatonin continued to have that beneficial effect on the offspring's fertility.

When nanoparticles and antioxidants were studied in combination, our data strongly suggest that AgNPs have the strongest detrimental impact on mobility and fertility. AuNPs and TiO2 also exert a detrimental effect, but not to the extent of AgNPs. NAC and quercetin were unable to reverse the detrimental effect in both mobility and fertility assays, even though they were able to do so when combined with TiO2 and AuNPs. Melatonin was the only antioxidant that exhibited a reversal effect on the negative impact of AgNPs.

Data was analyzed for statistical significance using one-way ANOVA of the trial averages of each replicate. Each replicate, using the mean across the trials, was assessed on its significance from the control groups using Tukey's post-hoc test. For analyzing the fertility assays, due to the single-time measurement nature of that assay, all raw data were utilized in the data analysis. This data analysis was conducted using a significance level (α) of 0.05.

The data analysis also brought irregularities in the parent generations' results forward, such as silver performing unusually high in the P1 mobility assay than what would be expected when compared to its results in the other assay. Error bars were calculated using n=5.

The results of this data analysis primarily supported our interpretation of the results. Individual conditions demonstrated significant differences from the control for all of the nanoparticles across both types of assays, while the positive effect of the antioxidants was primarily insignificant in the mobility assays, except for melatonin, which appeared statistically significant in both mobility assays. All chemicals tested were significant in at least one of the two fertility assays. Conversely to the individual conditions, the combined condition assays demonstrated little significant difference from the control in mobility, and none in either generation for fertility. The outliers in mobility were groups C (Silver + NAC), F (Silver + Quercetin), and G (TiO2 + Melatonin), as discussed earlier. These results verify our hypothesis and previous postulations as they indicate a clear reversal effect from the antioxidants.

Limitations and Further Experimentation:

These experiments were performed in an informal, at-home lab setting with experimental materials limited to commercially available resources. This experiment was conducted on a tight 7-week schedule and thus was limited to observing 2 generations of *Drosophila melanogaster*. While this setup allowed us to observe and measure the behavior of *Drosophila* flies, we were unable to perform any genetic or biochemical experiments to follow up on our observations. The lack of microgram-precise measuring tools also made it difficult to measure the exact concentrations of all of the vials. We improvised by reconstituting a concentrated solution before further serial dilution. Therefore, we need to keep in mind that there might be variations resulting from this technical issue. Finally, it is important to note that none of the tests are perfect measurements of the physical or reproductive health of *Drosophila melanogaster* and are subject to both human and experimental variation.

Nonetheless, our observations brought up some interesting questions for future experiments. Further experiments

DOI: 10.36838/v7i10.36

on this topic will include an expanded scope of nanoparticles and reversal compounds while also studying transgenerational impacts across more than 2 generations. If the scope of antioxidants tested were expanded, it would be interesting to test compounds reported to alleviate neurotoxic effects. Furthermore, our data raises interesting follow-up questions on whether *Drosophila* genetic variations would give flies an edge in survival: Are there any genetic mutations that would allow flies to resist nanoparticles' detrimental effect, or enhance the beneficial effect of antioxidants? Advancing this topic into gene analysis can give us an insight into genetic variation in flies, and the results are potentially translatable to human health. Finally, our data identified a few compounds with a significant impact on *Drosophila* health, such as silver nanoparticles and melatonin. Further studies can include biochemical assays to identify the exact cellular pathways through which these compounds exert their effect.

Conclusion

This study further verifies the significant impact that the tested nanoparticles exhibit on the mobility and fertility of Drosophila melanogaster, as well as the potential for mitigating these effects exhibited by the tested antioxidants. Gold, silver, and titanium dioxide nanoparticles exhibited detrimental effects on both fertility and mobility, with Gold and silver demonstrating transgenerational impacts. The antioxidants: melatonin, quercetin, and NAC all demonstrated a positive effect on these metrics, with melatonin demonstrating these effects transgenerationally.

Notably, the results of the second generation of flies suggest that treatment with the tested antioxidants can reverse transgenerational impacts in the nanoparticles tested, with melatonin being especially effective, as it was the only antioxidant to counteract the damaging effects of every nanoparticle tested. These findings highlight the potential for antioxidants to counteract the long-term biological consequences of nanoparticle exposure. However, considering the limitations of this study, including an informal lab setting, lack of genetic analysis, and restricted generational scope, further research is necessary. Future research might expand the range of nanoparticles tested, assess genetic influences, and expand the analysis to cover more generations. These studies would help to better understand the long-term implications of nanoparticle exposure.

In summary, this research has offered some valuable insights into the interplay between oxidative stress and neurotoxicity in the physical and reproductive health of *Drosophila melanogaster*, but given its informal nature, it is my personal opinion that further and more thorough research would be beneficial in investigating how to mitigate transgenerational toxicity in living organisms.

Acknowledgments

I am incredibly grateful for Dr. Yi Han Ng's mentorship and guidance as I began this research journey. Her insights were invaluable to shaping this paper and inspiring me to pursue further biological research. I would also like to thank both of my parents for their extremely essential feedback and unwavering support. Special thank you to Dr. Karen O'Brien for her direction and wisdom throughout the process. I attest that the ideas, graphics, and writing in this paper are entirely my own.

References

1. Ajdary, M.; Keyhanfar, F.; Moosavi, M. A.; Shabani, R.; Mehdizadeh, M.; Varma, R. S. Potential Toxicity of Nanoparticles on the Reproductive System Animal Models: A Review. *Journal of Reproductive Immunology* **2021**, *148*, 103384. https://doi.org/10.1016/j.jri.2021.103384.
2. Lihui, X.; Zhao, J.; Skonieczna, M.; Zhou, P.; Huang, R. Nanoparticles-Induced Potential Toxicity on Human Health: Applications, Toxicity Mechanisms, and Evaluation Models. *MedComm* **2023**, *4* (4). https://doi.org/10.1002/mco2.327.
3. Raj, A.; Shah, P.; Agrawal, N. Impact of Nanoparticles on Behavior and Physiology of *Drosophila Melanogaster*. *Springer eBooks* **2020**, *59*–67. https://doi.org/10.1007/978-981-15-5522-0_3.
4. Philbrook, N. A.; Winn, L. M.; A. R. M. Nabiul Afrooz; Saleh, N. B.; Walker, V. K. The Effect of TiO2 and Ag Nanoparticles on Reproduction and Development of Drosophila Melanogaster and CD-1 Mice. **2011**, *257* (3), 429–436. https://doi.org/10.1016/j.taap.2011.09.027.
5. Aloisi, M.; Rossi, G.; Colafarina, S.; Guido, M.; Cecconi, S.; Poma, A. The Impact of Metal Nanoparticles on Female Reproductive System: Risks and Opportunities. *International Journal of Environmental Research and Public Health* **2022**, *19* (21), 13748–13748. https://doi.org/10.3390/ijerph192113748.
6. Zhang, X.; Song, Y.; Wang, J.; Wu, C.; Xiang, H.; Hu, J.; Gong, H.; Sun, M. Chronic Exposure to Titanium Dioxide Nanoparticles Induces Deficits of Locomotor Behavior by Disrupting the Development of NMJ in Drosophila. *The Science of the total environment* **2023**, *888*, 164076. https://doi.org/10.1016/j.scitotenv.2023.164076.
7. Vecchio, G.; Galeone, A.; Brunetti, V.; Maiorano, G.; Sabella, S.; Cingolani, R.; Pier Paolo Pompa. Concentration-Dependent, Size-Independent Toxicity of Citrate-Capped AuNPs in Drosophila Melanogaster. *PLOS ONE* **2012**, *7* (1), e29980–e29980. https://doi.org/10.1371/journal.pone.0029980.
8. Rand, M. D.; Tennessen, J. M.; Trudy; Robert. Perspectives on the Drosophila Melanogaster Model for Advances in Toxicological Science. *Current protocols* **2023**, *3* (8). https://doi.org/10.1002/cpz1.870.
9. Jovanović, B.; Jovanović, N.; Cvetković, V. J.; Matić, S.; Stanić, S.; Whitley, E. M.; Mitrović, T. L. The Effects of a Human Food Additive, Titanium Dioxide Nanoparticles E171, on Drosophila Melanogaster - a 20 Generation Dietary Exposure Experiment. *Scientific Reports* **2018**, *8* (1), 17922. https://doi.org/10.1038/s41598-018-36174-w.
10. Zhang, X.; Song, Y.; Wang, J.; Wu, C.; Xiang, H.; Hu, J.; Gong, H.; Sun, M. Chronic Exposure to Titanium Dioxide Nanoparticles Induces Deficits of Locomotor Behavior by Disrupting the Development of NMJ in Drosophila. *The Science of the total environment* **2023**, *888*, 164076. https://doi.org/10.1016/j.scitotenv.2023.164076.
11. Rinaldi, C.; Carrero-Martinez, F.; Herrera, A.; Vega-Alvarez, S. Tissue-Specific Direct Microtransfer of Nanomaterials into Drosophila Embryos as a Versatile *in Vivo* Test Bed for Nanomaterial Toxicity Assessment. *International Journal of Nanomedicine* **2014**, *2031*. https://doi.org/10.2147/ijn.s56459.
12. Sadowska-Bartosz, I.; Bartosz, G. Effect of Antioxidants Supplementation on Aging and Longevity. *BioMed Research International* **2014**, *2014*, 1–17. https://doi.org/10.1155/2014/404680.
13. Zhu, A.; Zheng, F.; Zhang, W.; Li, L.; Li, Y.; Hu, H.; Wu, Y.; Bao, W.; Li, G.; Wang, Q.; Li, H. Oxidation and Antioxidation of Natural Products in the Model Organism Caenorhabditis Ele-

gans. *Antioxidants* **2022**, *11* (4), 705–705. https://doi.org/10.3390/antiox11040705.
14. Atkuri, K. R.; Mantovani, J. J.; Herzenberg, L. A.; Herzenberg, L. A. N-Acetylcysteine—a Safe Antidote for Cysteine/Glutathione Deficiency. *Current Opinion in Pharmacology* **2007**, *7* (4), 355–359. https://doi.org/10.1016/j.coph.2007.04.005.
15. Raghu, G.; Berk, M.; Campochiaro, P. A.; Jaeschke, H.; Marenzi, G.; Richeldi, L.; Wen, F.-Q.; Nicoletti, F.; Calverley, P. M. A. The Multifaceted Therapeutic Role of N-Acetylcysteine (NAC) in Disorders Characterized by Oxidative Stress. *Current Neuropharmacology* **2021**, *19* (8), 1202–1224. https://doi.org/10.2174/1570159x19666201230144109.
16. Nalika, N.; Waseem, M.; Kaushik, P.; Salman, M.; Andrabi, S. S.; Jamal, A.; Parvez, S. Role of Melatonin and Quercetin as Countermeasures to the Mitochondrial Dysfunction Induced by Titanium Dioxide Nanoparticles. *Life Sciences* **2023**, *328*, 121403. https://doi.org/10.1016/j.lfs.2023.121403.
17. J. Reiter, R.; C. Manchester, L.; Tan, D.-X. Neurotoxins: Free Radical Mechanisms and Melatonin Protection. *Current Neuropharmacology* **2010**, *8* (3), 194–210. https://doi.org/10.2174/157015910792246236.
18. Reiter, R. J.; Mayo, J. C.; Tan, D.-X.; Sainz, R. M.; Alatorre-Jimenez, M.; Qin, L. Melatonin as an Antioxidant: Under Promises but over Delivers. *Journal of Pineal Research* **2016**, *61* (3), 253–278. https://doi.org/10.1111/jpi.12360.
19. Qi, W.; Qi, W.; Xiong, D.; Long, M. Quercetin: Its Antioxidant Mechanism, Antibacterial Properties and Potential Application in Prevention and Control of Toxipathy. *Molecules* **2022**, *27* (19), 6545. https://doi.org/10.3390/molecules27196545.
20. Xuan, L.; Ju, Z.; Skonieczna, M.; Zhou, P.; Huang, R. Nanoparticles-induced potential toxicity on human health: Applications, toxicity mechanisms, and evaluation models. *MedComm* **2023**, *4* (4). https://doi.org/10.1002/mco2.327.
21. Pompa, P. P.; Vecchio, G.; Galeone, A.; Brunetti, V.; Sabella, S.; Maiorano, G.; Falqui, A.; Bertoni, G.; Cingolani, R. In Vivo toxicity assessment of gold nanoparticles in *Drosophila melanogaster*. *Nano Research* **2011**, *4* (4), 405–413. https://doi.org/10.1007/s12274-011-0095-z.
22. Basak, A. K.; Chatterjee, T.; Chakravarty, A.; Ghosh, S. K. Silver nanoparticle–induced developmental inhibition of Drosophila melanogaster accompanies disruption of genetic material of larval neural stem cells and non-neuronal cells. *Environmental Monitoring and Assessment* **2019**, *191* (8). https://doi.org/10.1007/s10661-019-7630-x.
23. Millet-Boureima, C.; Ennis, C. C.; Jamison, J.; McSweeney, S.; Park, A.; Gamberi, C. Empowering Melatonin Therapeutics with Drosophila Models. *Diseases* **2021**, *9* (4), 67. https://doi.org/10.3390/diseases9040067.
24. Brack, C.; Bechter-Thüring, E.; Labuhn, M. N-Acetylcysteine slows down ageing and increases the life span of Drosophila melanogaster. *Cellular and Molecular Life Sciences* **1997**, *53* (11–12), 960–966. https://doi.org/10.1007/pl00013199.
25. Omoboyowa, D. A.; Olugbenga, S. T.; Adetuyi, F. D.; Akinsulure, S. T.; Akinwande, K. M.; Iwuji, C. B.; Ayekolu, S. F. Modulatory effects of selected compounds on oxidative stress in hydrogen peroxide-induced *Drosophila melanogaster*. *Pharmacological Research - Modern Chinese Medicine* **2022**, *5*, 100169. https://doi.org/10.1016/j.prmcm.2022.100169.

Author

Andrew (Drew) Mozinski is a junior at Crimson Global Academy. He aspires to study biochemistry in college and hopes to pursue a profession in scientific research. He is also a passionate rower who is looking forward to competing in the US Rowing Youth National Championship 1x this spring.

DOI: 10.36838/v7i10.36

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Evaluating the Impact of the Completion of Beijing Subway Line 6 on House Prices and Its Implications for Urban Planning

Eric Tang

Chinese International School, 1 Hau Yuen Path, Braemar Hill, Hong Kong, China; etang.hkus@gmail.com

ABSTRACT: The paper examines the impact of the completion of a new urban rail transit (URT), Line 6, on nearby community housing prices between 2011-2014 in Beijing, well ahead of the Chinese government's repression of speculative housing investments and the fall of China's newborn in 2016, the government restriction on developers' leverage ratio in 2020, and the structural decline of China's total population in 2022. It estimates empirically the pulling effect of the completion of a new urban rail transit on the average housing prices of the communities along the route. The conclusion, with the integration of the hedonic pricing method (HPM) and the difference in difference (DiD) method, is that the completion of subway Line 6 has a significant impact on nearby communities' housing prices. In addition, the difference between walkable and non-walkable distances is evaluated, and the clustering effect is refuted. Finally, the difference in impact among urban, semi-urban, and suburban areas is quantified with the finding that the completion of a new subway has the most significant impact on the communities in semi-urban areas. Based on the findings, advice in urban planning for other big cities with large populations in fast-growing economies is provided.

KEYWORDS: Behavioral and Social Sciences, Sociology and Social Psychology, Urban Rail Transit (URT), Housing Price, Urban Planning.

Introduction

Although it seems apparent that the addition of a public transport line will increase the land value along its route, given residents' reduced cost of commute and improved accessibility to the central business district (CBD), it is critical to differentiate between correlation and causality. Whether the completion of a new subway fosters more economic activities and then causes higher housing prices or economic development is the key driver behind the subway construction and housing price appreciation remains unclear. In our examination, we try to identify and quantify the cause and the result, and we intentionally exclude the first and last stations of the subway line that we examine to remove the excess distortion due to speculations on government policy announcements.

Our particular interest in this topic resides in its quantitative evaluation of the impact of public transportation investment on the urban economy. As of the end of 2024, there are 50 cities in the world that have a population of above 7 million, about 2/3 (34 cities) of these cities are in Asia, among which 56% (19 cities) are outside of China. With the fast population growth in Southeast Asia and the further urbanization of rural migrants in these economies, lessons learned from China's fast urbanization period are relevant. Where and how a limited fiscal budget should be invested to maximize the outcome and benefit most people are puzzling questions facing all policy makers and urban planners. Our research provides a reference for decision makers on these issues.

Another reason for our study is the limited number of comparable studies for Mainland China, particularly during the specific period examined, right after the completion of a new subway line. There are, however, plenty of empirical analyses done across many developed countries. Just to name a few, Baldassare[1] examined the impact of transit rail on the nearby land value based on social elements and environmental attributes, while Prior[2] and Riley[3] investigated property value along the new Joban line in Tokyo and the London Metro Jubilee extension, respectively.

Among the China-focused research, Xiao, Webster, and Orford[4] explored the linkages between urban configuration (including the subway network) and micro-level house price movement, taking the case of the city of Nanjing. The paper employed a spatial-network analysis method to track changes in transport accessibility and implied a generally positive relationship between accessibility and property prices and a negative relationship where spillover effects led to new congestion hot spots. Tan, Zhao & Li[5] found that the opening of new subway stations has led to significant increases in subway usage, trip duration, and trip distance within a 2000-m radius, and that commuting trips have been more affected than home-based non-commuting trips.

Lai[6] investigated the factors influencing housing prices in Beijing using a multiple linear regression model (MLR) and found that subway accessibility positively correlates with housing prices, while increased square footage is negatively associated with price. The paper suggests that smaller homes with better subway access are more valuable. Wang[7] discussed how urban rail transit has influenced China's urban spatial structure and real estate supply, and produced externalities on real estate value. The impact of rail transit on housing prices shows a trend of first increasing and then decreasing, depending on the distance from the subway stations.

DOI: 10.36838/v7i10.45

In our study, we combined the Hedonic Price Model (HPM) and Difference in Difference (DiD) method to capture the price changes in the housing market along the new subway line, with particular comparisons between the prior and post-completion period, which will be discussed in more detail in the later sections.

The Beijing Subway started operations in 1969 and is a rapid transit rail network that serves the urban and suburban districts of Beijing. As of 2024, the network has 29 lines, 522 stations, and 879km of track in operation. Beijing's population increased 2.2 million from 19.6 million in 2010 to 21.8 million in 2015, and has been range-bound between 21.5-22.2 million thereafter. As the capital of China with ~3000 years of history and the best education and healthcare resources in the country, the stagnation of its population growth was a result of government intentional control since 2015 and the roll out of the construction plan of the Xiongan city, a completely new city 120 km from Beijing, to help promote regional development, disperse Beijing's non-capital function, and ease the pressure of Beijing.

The average housing price in Beijing was close to RMB 20,000/sqm in 2011, RMB 30,000/sqm in 2015, RMB 40,000/sqm in 2020, and RMB 45,000/sqm in 2024. The period we examined, between 2011 and 2014, was less affected by speculative investments, government-engineered lending crackdowns, or population control. By the end of 2014, the Beijing subway network had 23 lines, 340 stations, and 527 km of track in operation, and was one of the most loaded in the world at 3.4 billion trips per year, averaging 9.3 million per day. The existing network still could not sufficiently meet the city's mass transportation needs, and the government funded expansion planned for a 1/3 of extension in subway track length by 2015 and another 1/3 by 2020.

Beijing Subway Line 6 is a rapid transit rail line connecting the west and the east of the city, running north of and in parallel to Line 1, the oldest subway line in Beijing, to ease the congestion of the latter. Stage I of Line 6 extends between Wuluju Station in Haidian District in the west to the Caofang Station in Tongzhou District in the east, spanning 20 stations and 31km. It officially started operation on December 30, 2012. Stage II extends to the east, further from Caofang station to Lucheng station in Tongzhou District, adding 8 stations and 12km. It officially started operation on December 28, 2014. Stage III extends to the west from Wuluju station to Jinanqiao station in Shijingshan District, adding 6 stations and 10.6km, and started operation on December 30, 2018, while Stage IV extends further to the east to Pinggu District and started operation on December 31, 2021. As of the end of 2024, Line 6 has 35 stations in total and spans 53km. It is one of the key subway lines in Beijing. Six additional lines were built in Beijing after 2014.

The period we examined covers the completion of Stage I and the ramp-up of Stage II. The focus of our research is the causal effect of the completion of a new subway line on economic development. Instead of macro variables such as GDP per capita or GDP, we chose micro factors contributing to house prices as proxies for the impact of government infrastructure investments. There are three main reasons for doing so. First, GDP is an assessment of overall economic output; it cannot reflect in detail the segmented economic development driven by a particular subway line completion. Second, the Tiebout[8] model on local public finance claims that the quality of local public service, including public transportation, is capitalized into the housing prices through households' residential location decisions. Third, the housing prices we use are supported by market data and are more reliable.

Theoretical analysis

Travel costs, as measured by availability and convenience of transport, transportation expenses, and travel time, are the key elements that urban transit affects house prices. R. F. Muth[9] developed a theoretical framework to understand how households maximize their utility given their income and transportation costs. His framework explains the spatial distribution of population and housing within urban areas and shows that the location of a dwelling unit depends on the trade-off between transportation expenses and housing expenses.

W. Bruce Allen's study[10] examines the impact of the New Jersey Transit System on housing prices in the Philadelphia metropolitan area. Using the hedonic pricing model, his study finds that housing prices increase significantly with proximity to transit stations. Specifically, it reveals that for each dollar reduction in transportation expenses, housing values increase by approximately $149 (in 1971 dollars). The increase is attributed to the time savings and convenience provided by the transit system.

So, Tse & Ganesan[11] discovered in their study of the Hong Kong housing market that increased transportation facility enhances access to employment, and help stimulate economic activities through the large stream of people brought by the subway, forming new shopping complexes and offering extra job opportunities. The model includes structural, physical, and environmental attributes, and reveals that the accessibility of public transportation is highly valued by residents and is reflected in higher housing prices.

Another important factor is the location of the dwelling unit in the city, whether it is in the Central Business District (CBD) or a suburban area. Li H, Wei Y, Wu Y, & Tian G[12] examined the spatial variations of housing prices in Shanghai, focusing on the relationship between housing prices and distance to the CBD. The results show that housing prices generally decay with an increase in distance from the CBD. The study also highlights the impact of accessibility and service amenities on housing prices.

Stegman,[13] who researched the construction of a railway in San Francisco on the housing price, found that properties closer to Bay Area Rapid Transit (BART) stations experienced significant increases in value, reflecting the capitalization of reduced travel costs and improved accessibility. Separately, Rena Sivitanidou[14] pointed out that the degree to which the Rail Transit in Los Angeles affects the commercial property value varies from major CBD to minor CBD.

Other mechanisms of the subway's impact on house prices involve commercial speculation and irrational expectations. This is most obvious in the case where housing price experiences an abrupt increase upon the announcement of a new subway line passing through the neighborhood. DeFusco A, Nathanson C, & Zwick E.[15] finds, from 50 million home sales during the last U.S. housing cycle, that cities with larger speculative booms have sharper increases in unsold listings as the market turns, and more severe busts.

Model and Data

Baseline Model:

In our baseline model, we combine the HPM and the DiD method.

Hedonic price analysis is a technique that studies the demand side of housing, assuming that a property is sold as a package of inherent attributes (Rosen[16]). Hedonic prices are, in fact, the implicit valuations of the characteristics of the housing unit, such as quality, location, etc., which influence the market price of the house, and the implicit prices can be estimated by regression analysis. The model is particularly useful in understanding how consumers value different attributes and how these values are reflected in market prices.

The Difference in Difference method is a statistical technique used to evaluate the causal effect of a treatment or intervention by comparing the changes in outcomes over time between a treatment group and a control group. The method involves calculating the difference in outcomes between the treatment and control groups before and after the intervention. The basic idea is to isolate the effect of the treatment by comparing the changes in outcomes for the treatment group relative to the control group.

Based on the hedonic hypothesis, we first pinpoint the characteristics that significantly affect the average house price in a specific community, and then compare the difference in the house prices between the control group and the treatment group, before and after the subway's operation. In the baseline model, the treatment group includes all the communities near the subway station (within 1.5 km), and the control group consists of randomly picked communities, which are far from the subway Line 6 and do not have any newly established subway stations nearby during the time period.

Our baseline model is described as follows:

$$\log(P_{it}) = \beta_0 + \beta_1 Treated_i * Post_t + \beta_2 Post_t + \delta Z_i + \varepsilon_{it}$$

P_{it} is the dependent variable, and the average house price (per square meter) of community i in year t. The key independent variables include: 1) $Treated_i$, the dummy variable for the treatment group: 1 for the communities near the subway station and 0 for those far away and unlikely to be influenced by the subway line. 2) $Post_t$, the dummy for the subway's operation: 1 if later than December 30, 2012, and 0 if earlier. 3) Z_i, a vector of community characteristics attributing to the house price, among which a) Age_{it}, is measured by the years from the establishment of community i to year t; b) Bus_i, is the total number of bus lines within 1 kilometer; c) $CBDDist_i$, is the Euclidean distance from the community to the CBD; d) $Floor_i$ is the average floor numbers; e) D_{1i}, D_{2i}, D_{3i}, D_{4i}, are the dummy for 4 house types, tower-type, slab-type, bungalow and courtyard; f) $School_i$ is the dummy for school district; g) $Trans_i$, is the dummy for transfer station: 1 if the nearest subway station is a transfer station and 0 if not. Lastly, ε_{it} is the error term.

The data employed in this research were obtained from established real estate agent companies and websites. The housing attributes were collected through on-site observations and surveys. The sample size is 422 communities, over 4 years from 2011 to 2014.

The model measures the effect of the completion of the new subway Line 6 on the treated communities. β_1 measures the difference between the controlled group and the treatment group. β_2 measures the common trend of house prices among all the communities during the period; vector δ measures the house characteristics' impact on house prices.

Main caveats:

We highlight two main caveats of our baseline model.

First, the factors we examine are not complete: other hedonic pricing models could also incorporate additional characteristics that are positively correlated with house prices, such as the garage size, parking area, hospitals nearby, etc. As a result, the OLS (Ordinary Least Squares) estimators can be up-biased. However, during the first two decades of the 21st century, particularly in Beijing, location factors were extraordinarily significant in housing prices, and the other living conditions' impacts were subdued. In fact, since the commercial housing reform in 1998 in China, the apartments built during a certain period mostly have similar features and living conditions, thus most of the omitted attributes related to the community's quality can be reflected by the house age. Given that it is impossible to include all the characteristics in our model, we interpret our estimates as the upper-bound effect of the subway on house prices.

Second, the most likely limitation on the Hedonic pricing model is multicollinearity. Multicollinearity refers to a situation where two or more independent variables in a regression model are highly correlated with each other. This can lead to unstable and unreliable estimates of the regression coefficients, distort the estimated implicit prices of housing attributes, and make it difficult to accurately assess the impact. As Butler[17] comments, one is unlikely to find a fourteen-bedroom property with only one bathroom and vice versa! However, whether multicollinearity is an issue in concluding any given data set can only be tested statistically.

Results and Discussion

A. Descriptive Statistics:

Tables 1 and 2 display the descriptive statistics of all the variables used in our empirical analysis, listing the descriptive results of the communities along the subway line (treatment group in the baseline model) and off the subway line (control group), respectively.

DOI: 10.36838/v7i10.45

Table 1: Descriptive statistics of key variables (along subway communities). This table presents the summary statistics of the key characteristics of the communities along Beijing subway line 6. House Price 2011-2014 are the average house prices (RMB per square meter) of each community in each year. The distance to Subway is walking distance from the community center to the nearest subway station, measured by Google Maps in meters. Distance to CBD is the Euclidean distance from the community to Beijing's Central Business District (CBD) area measured in kilometers. House Type is a four-indicator variable, equal to one if the main building type of the community is tower, slab, bungalow, or courtyard, respectively, and zero if otherwise. Year of Build is the year when the community was built. Floor Number is the average number of floors across all the buildings in the community. The sample includes all communities along subway line 6 provided by the real estate agent companies. The data span from 2011 to 2014.

Variable	Obs	Mean	Std. Dev.	Min	Max
House Price 2011 (RMB/sqm)	170	28447	8548	10406	51786
House Price 2012 (RMB/sqm)	338	32021	10306	12145	64652
House Price 2013 (RMB/sqm)	354	42635	13230	15808	89277
House Price 2014 (RMB/sqm)	340	43719	12103	19687	98039
Distance to Subway (m)	420	755	316	4	1626
Distance to CBD (km)	422	10.5	4.6	3.5	23.8
House Type					
Dummy1 for Tower-type	421	0.22	0.41	0	1
Dummy2 for Slab-type	419	0.47	0.50	0	1
Dummy3 for Bungalow	421	0.02	0.14	0	1
Dummy4 for Courtyard	421	0.01	0.10	0	1
Year of Build	421	1995	10	1960	2014
Floor Number	421	13	8	0	57
Dummy for School District	421	0.32	0.47	0	1
Number of Bus Lines	422	18	7	5	29
Dummy for Transfer Station	422	0.54	0.50	0	1

As shown in the exhibits, the average housing prices kept rising for both groups, but the treatment group's price rose faster. The average house price of the control group in Table 2 rose by 5.2%, 10.3%, and 9.6% in 2012, 2013, and 2014, respectively. It is in line with the common trend of house prices in Beijing. However, the treatment group's price in Table 1 rose sharply from 2012 to 2013, specifically 33.1%, as the subway began to operate on December 30, 2012. This timing aligns closely with the theoretical expectation outlined in our analysis. The housing price growth rate in the treatment group slowed down dramatically from 2013 to 2014, to approximately 2.5%, even slower than the common trend. It is probably due to buyers' overreaction in the previous year, so there likely exists a mean reversion process.

Table 2: Descriptive statistics of key variables (off subway communities). This table presents the summary statistics of the key characteristics of the communities in the control group, which are located off the subway line 6. Variable definitions are the same as Table 1.

Variable	Obs	Mean	Std. Dev.	Min	Max
House Price 2011 (RMB/sqm)	20	23682	5425	15040	38745
House Price 2012 (RMB/sqm)	20	24917	4540	18498	36485
House Price 2013 (RMB/sqm)	20	27491	5201	17640	38485
House Price 2014 (RMB/sqm)	20	30134	5534	20192	46582
Distance to CBD (km)	20	12.7	3.8	7.7	21.8
House Type					
Dummyl for Tower-type	20	0.20	0.41	0	1
Dummy2 for Slab-type	20	0.25	0.44	0	1
Dummy3 for Bungalow	20	0.00	0.00	0	0
Dummy4 for Courtyard	20	0.00	0.00	0	0
Year of Build	20	1998	6	1980	2007
Floor Number	20	18	5	6	25
Dummy for School District	20	0.45	0.51	0	1
Number of Bus Lines	20	27	15	6	68
Dummy for Transfer Station	20	0	0	0	0

B. Pulling Effect on House Prices:

The "pulling effect" of subways on urban housing prices refers to the phenomenon where the construction and operation of subway lines significantly increase the market value of nearby residential properties. This effect is driven by several key factors, including improved accessibility, reduced commuting costs, and enhanced neighborhood desirability. We first examine the completion of Line 6's impact on average house prices. Table 3 shows the time series regression result of our baseline model.

As shown in Table 3, columns 1-3, we apply the model using the data of 2012 and 2013, 2011 and 2014, and 2011-2014, respectively, to balance between the pre-treatment and post-treatment periods. Not surprisingly, we find positive and significant (1% level) coefficients of the Post*Treated term, which indicates that the completion of the new subway line causes the house price to go up by 40.1% from 2012 to 2013, and 34.2% from 2011 to 2014. The immediate price jump was mainly due to buyers' irrational expectations and herding behavior, while the milder long-run appreciation mainly resulted from the improved convenience.

From the hedonic pricing model statistics, we conclude that, when determining the house prices, all the following characteristics play a significantly important role: distance to CBD, house age, whether in a school district, whether near a transfer station, floor number, and the number of bus lines around. With regard to the house type, whether it is tower-type, slab-type, or bungalow, does not matter, but given all other factors are the same, courtyard houses have significantly higher prices (around 70% higher).

Table 3: OLS estimation of the completion of the new subway on housing prices based on time series. We find positive and significant (1% level) coefficients of the Post*Treated term, which indicates that the completion of the new subway line causes the house price to go up by 40.1% from 2012 to 2013, and 34.2% from 2011 to 2014.

Dependent Variable	Log (Average House Price)		
Sample:	All Communities; Year: 2011-2014		
Years Selected	2012, 2013	2011, 2014	2011-2014
	(1)	(2)	(3)
Post*Treated	0.401***	0.342***	0.367***
	(0.0608)	(0.0524)	(0.0410)
Post	-0.100	0.105**	-0.010
	(0.0605)	(0.0527)	(0.0409)
Distance to CBD	-0.0333***	-0.0289***	-0.0317***
	(0.0027)	(0.0027)	(0.0020)
Age	-0.00390***	-0.00475***	-0.00397***
	(0.0013)	(0.0013)	(0.0009)
House Type			
Dummyl for Tower-type	-0.04	-0.01	-0.03
	(0.0281)	(0.0282)	(0.0203)
Dummy2 for Slab-type	0.03	0.02	0.02
	(0.0253)	(0.0246)	(0.0180)
Dummy3 for Bungalow	0.21	0.15	0.2
	(0.1820)	(0.2170)	(0.1400)
Dummy4 for Courtyard	0.631**	0.711***	0.665***
	(0.2550)	(0.2160)	(0.1710)
Floor Number	-0.00324**	-0.00329**	-0.00311***
	(0.0015)	(0.0016)	(0.0011)
Dummy for School District	0.222***	0.206***	0.213***
	(0.0213)	(0.0206)	(0.0151)
Number of Buses	0.00798***	0.00639***	0.00699***
	(0.0014)	(0.0014)	(0.0010)
Dummy for Transfer Station	0.0511**	0.0481**	0.0516***
	(0.0223)	(0.0222)	(0.0160)
Constant	10.53***	10.43***	10.50***
	(0.0626)	(0.0623)	(0.0450)
Observations	731	548	1279
R-squared	0.545	0.663	0.584

Notes: Standard errors in parentheses ***p<0.01, **p<0.05, *p<0.1

C. Clustering Effect vs. Spillover Effect:

The clustering effect of urban rail transit refers to the tendency of economic activities, population, and development to concentrate around rail transit stations. The spillover effect of urban rail transit refers to the unintended consequences or impacts that the URT system has on surrounding areas, including land use, economic activity, and environmental conditions. These effects can be both positive and negative, influencing property values, urban development, and overall economic growth.

The common assumption is that the completion of a new subway line can mostly affect the house price of the communities within the walkable distance (0.8 km) of the subway stations, because most people in the nearby communities tend to walk to the stations instead of driving there, while people owning private cars are unlikely to take the subway. People could also take buses to the subway stations, but those taking buses would then not live within 1.5 km of the stations.

As the supporter of this perspective, Heenan[18] investigated the clustering effect of urban rail transit, stating that as the distance increases, the houses beyond the walkable range may even experience a price decline, because buyers prefer houses within walkable distance, and the demand for outer houses will drop. The study discusses how rapid transit systems can lead to more efficient urban growth patterns and encourages development in areas with easy access to transit. Empirically, Cockerill and Stanley[19] confirmed Heenan's theory that the urban rail transits only have a significant influence on the price of the houses within 0.8 km of the stations.

Table 4: OLS estimation of clustering/spillover effect based on distance to subway station. The Post terms indicating the common trend are all significant at the 1% level, which implies that the completion of the new subway line has a very positive spillover effect on many nearby communities, even beyond the walkable distance.

Dependent Variable	**Log (Average House Price)**				
Sample:	Communities along the subway line; Year: 2011-2014				
Distance to Subway Station	0-200	0-400	0-600	0-800	0-1000
	(1)	(2)	(3)	(4)	(5)
Post*Treated200	0.0237				
	(0.0443)				
Post*Treated400		0.022			
		(0.0227)			
Post*Treated600			0.0095		
			(0.0194)		
Post*Treated800				-0.0095	
				(0.0183)	
Post*Treated1000					-0.0074
					(0.0214)
Post	0.340***	0.337***	0.338***	0.347***	0.347***
	(0.0141)	(0.0146)	(0.0153)	(0.0171)	(0.0215)
Distance to CBD	-0.0312***	-0.0312***	-0.0312***	-0.0311***	-0.0312***
	(0.0019)	(0.0019)	(0.0019)	(0.0019)	(0.0019)
Age	-0.00401***	-0.00398***	-0.00397***	-0.00401***	-0.00401***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
House Type					
Dummyl for Tower-type	-0.034	-0.0336	-0.0334	-0.0348	-0.0343
	(0.0211)	(0.0211)	(0.0211)	(0.0212)	(0.0211)
Dummy2 for Slab-type	0.0269	0.0269	0.0268	0.0262	0.0266
	(0.0189)	(0.0189)	(0.0189)	(0.0189)	(0.0189)
Dummy3 for Bungalow	0.239*	0.234*	0.237*	0.240*	0.239*
	(0.1390)	(0.1390)	(0.1390)	(0.1390)	(0.1390)
Dummy4 for Courtyard	0.705***	0.708***	0.707***	0.698***	0.698***
	(0.1690)	(0.1690)	(0.1690)	(0.1700)	(0.1700)
Floor Number	-0.00263**	-0.00261**	-0.00261**	-0.00262**	-0.00260**
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0012)
Dummy for School District	0.217***	0.217***	0.217***	0.216***	0.216***
	(0.0155)	(0.0155)	(0.0155)	(0.0155)	(0.0155)
Number of Bus Lines	0.00979***	0.00980***	0.00975***	0.00970***	0.00972***
	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)
Dummy for Transfer Station	0.0345***	0.0349***	0.0349***	0.0353***	0.0350***
	(0.0124)	(0.0124)	(0.0124)	(0.0124)	(0.0124)
Constant	10.45***	10.45***	10.46***	10.46***	10.46***
	(0.0466)	(0.0466)	(0.0467)	(0.0466)	(0.0466)
Observations	1199	1199	1199	1199	1199
R-squared	0.593	0.593	0.593	0.593	0.593

Notes: Standard errors in parentheses ***p<0.01, **p<0.05, *p<0.1

For this particular analysis, we apply a discrete treatment method to the communities along the new subway line and use the baseline model again. The treatment groups are <200m, <400m, <600m, <800m, and <1000m, respectively, in Table 4, columns 1-5. We try to examine the difference in price change between the treatment group and the control group. However, our empirical results contrast with the common-sense clustering effect.

The Post*Treated term in these five regression results is not statistically significant, which means there is not much difference in the price change between the walkable and non-walkable distances. The Post terms indicating the common trend, however, are all significant at the 1% level, which implies that the completion of the new subway line has a very positive spillover effect on many nearby communities, even beyond the walkable distance. The spillover effect may come from the increased productivity and the prosperity of commerce in the communities.

D. Distance to CBD and Decentralization Effect:

We also examine the completion of the new subway line in urban areas, semi-urban areas, and suburban areas, defined by their distance to the Central Business District. In particular, in columns 1-5 of Table 5, we run the regression for the communities whose distances to the CBD are: <5km, 5-10km, 10-15km, 15-20km, and >20km, respectively.

Table 5: OLS estimation of the completion of the subway line's effect on house prices based on distance to the CBD. In urban areas (< 5km to CBD), the completion of a new subway line's pulling effect on house prices is the lowest (19.3%). In a semi-urban area (5-15km to CBD), the effect is most significant (42.1% for 5-10 km and 34.8% for 10-15 km). In a suburban area (>15 km), the effect falls again (20.4% for 15-20 km and 17.1% for >20 km).

Dependent Variable	**Log (Average House Price)**				
Sample:	Communities along the subway line; Year: 2011-2014				
Distance to CBD	<5km	5km-10km	10km-15km	15km-20km	>20km
	(1)	(2)	(3)	(4)	(5)
Post*Treated	0.193***	0.421***	0.348***	0.204***	0.171***
	(0.0555)	(0.0423)	(0.0451)	(0.0452)	(0.0580)
Post	0.340***	-0.0319	0.0088	0.175***	0.192***
	(0.0141)	(0.0404)	(0.0435)	(0.0410)	(0.0420)
Distance to CBD	-0.0138**	-0.0218***	0.0116*	0.0045	-0.0246***
	(0.0060)	(0.0042)	(0.0067)	(0.0061)	(0.0060)
Age	0.0147***	-0.00432***	-0.0003	-0.0115***	-0.0029
	(0.0031)	(0.0011)	(0.0017)	(0.0024)	(0.0036)
House Type					
Dummyl for Tower-type	-0.0051	0.0528**	-0.0452	-0.0760*	-0.0122
	(0.0495)	(0.0264)	(0.0309)	(0.0420)	(0.0556)
Dummy2 for Slab-type	0.146***	0.121***	0.0028	-0.0624**	0.053
	(0.0355)	(0.0225)	(0.0313)	(0.0306)	(0.0489)
Dummy3 for Bungalow	0.415***				
	(0.1030)				
Dummy4 for Courtyard	0.730***				
	(0.1190)				
Floor Number	0.0169***	0.0011	-0.00395**	-0.0149***	0.0158***
	(0.0039)	(0.0014)	(0.0019)	(0.0026)	(0.0038)
Dummy for School District	0.0317	0.160***	0.223***	0.0748***	0.128***
	(0.0333)	(0.0186)	(0.0272)	(0.0280)	(0.0394)

DOI: 10.36838/v7i10.45

Number of Bus Lines	-0.00460**	-0.0001	0.0005	0.0158***	0.0003
	(0.0021)	(0.0013)	(0.0017)	(0.0013)	(0.0023)
Dummy for Transfer Station	0.307***	0.0298	0.0384	0.416***	
	(0.0827)	(0.0197)	(0.0264)	(0.0391)	
Constant	9.817***	10.50***	10.07***	10.05***	10.05***
	(0.1140)	(0.0620)	(0.1070)	(0.1240)	(0.1380)
Observations	135	565	446	276	148
R-squared	0.862	0.573	0.445	0.658	0.601

Notes: Standard errors in parentheses ***p<0.01,**p<0.05,*p<0.1

As shown in Table 5, in urban areas (< 5km to CBD), the completion of a new subway line's pulling effect on house prices is the lowest (19.3%). In a semi-urban area (5-15km to CBD), the effect is most significant (42.1% for 5-10 km and 34.8% for 10-15 km). In a suburban area (>15 km), the effect falls again (20.4% for 15-20 km and 17.1% for >20 km).

For urban areas, the transportation system, as well as other value-added amenities, is already well developed; thus, adding a new subway line does not have much impact. For suburb, residents' purchasing power is lower, and buyers' expectations for price appreciation are modest. In a semi-urban area, wage earners' residential density is extremely high. As such, people's demand for a cheap and convenient transportation system is correspondingly high. Thus, the completion of a new subway line has the most significant spillover effect on semi-urban areas.

Robustness Test

First, we pointed out earlier that multicollinearity is an unavoidable common problem in the hedonic pricing model. We test for collinearity of our model using the variance inflation factor (VIF) method, which measures how much the variance of an estimated regression coefficient is increased due to multicollinearity. A VIF value greater than 10 is generally considered indicative of a multicollinearity problem. Table 6 shows the test result of VIF, a range from 1.02 to 1.9 (less than 10), which suggests that multicollinearity is not a significant issue in our case.

Table 6: Multicollinearity test - VIF method. This table examines whether the variables used in the regression analysis exhibit multicollinearity using the Variance Inflation Factor (VIF) test. The VIF of each variable is equal to the inverse of (1 - R-squared), where R-squared is the coefficient of determination when regressing the variable on all other variables. A VIF value of 1 demonstrates no multicollinearity, above 5 or 10 suggests a multicollinearity problem that leads to unreliable regression output. A range from 1.02 to 1.90 suggests that multicollinearity is not a significant issue in our case.

Variable	VIF	1/VIF
Post	1.02	0.9843
Distance to Nearest Subway Station	1.03	0.9678
Distance to CBD	1.84	0.5421
Age	1.66	0.6015
Dummyl for Tower-type	1.68	0.5949
Dummy2 for Slab-type	1.90	0.5256
Dummy3 for Bungalow	1.04	0.9596
Dummy4 for Courtyard	1.03	0.9732
Floor Number	1.69	0.5906
Dummy for School District	1.19	0.8394
Number of Buses	1.34	0.7457
Dummy for Transfer Station	1.42	0.7029

Mean VIF: 1.4

Next, we conduct a homoscedasticity check of our model using the Cook-Weisberg test. Homoscedasticity is an assumption in linear regression models that the variance of the error terms is constant across all levels of the independent variables. This assumption is crucial for the validity and reliability of regression analysis. Violations of homoscedasticity, known as heteroscedasticity, can lead to biased standard errors and unreliable statistical inferences. The Cook-Weisberg test is a statistical test used to detect heteroscedasticity in a linear regression model. The results show that our regressions are all heteroscedastic.

We repeated our regressions using the White[20] heteroscedasticity correction, with the results shown in Table 7. The White heteroscedasticity correction is a method used to adjust the standard errors of the regression coefficients in the presence of heteroscedasticity. This correction ensures that the standard errors are robust to the presence of non-constant variance in the error terms, thereby providing more reliable statistical inferences. By comparing Table 3 and Table 7, we can draw similar conclusions. Therefore, despite the heteroscedasticity, the analysis results still hold.

Table 7: Heteroscedasticity correction. This table presents the same regression output as in Table 3 but relaxes the homoskedasticity assumption. To address the fact that residual terms in the regression are potentially heterogeneous across communities, we use the heteroskedasticity-robust standard errors developed by White to test the statistical significance of regression coefficients. The standard errors are reported in parentheses. ***p<0.01, **p<0.05, *p<0.1. By comparing Table 3 and Table 7, we can draw similar conclusions. Therefore, despite the heteroscedasticity, the analysis results still hold.

Dependent Variable	**Log (Average House Price)**		
Sample:	All Communities; Year: 2011-2014		
Years Selected	2012, 2013	2011, 2014	2011-2014
	(1)	(2)	(3)
Post*Treated	0.401***	0.342***	0.367***
	(0.0500)	(0.0445)	(0.0345)
Post	-0.0980*	0.105**	-0.010
	(0.0506)	(0.0462)	(0.0353)
Distance to CBD	-0.0333***	-0.0289***	-0.0317***
	(0.0026)	(0.0026)	(0.0019)
Age	-0.00390***	-0.00475***	-0.00397***
	(0.0013)	(0.0013)	(0.0009)
House Type			
Dummyl for Tower-type	-0.036	-0.009	-0.027
	(0.0293)	(0.0288)	(0.0211)
Dummy2 for Slab-type	0.028	0.020	0.024
	(0.0261)	(0.0248)	(0.0185)
Dummy3 for Bungalow	0.214	0.153***	0.204
	(0.2550)	(0.0410)	(0.1580)
Dummy4 for Courtyard	0.631***	0.711***	0.665***
	(0.0381)	(0.0367)	(0.0439)
Floor Number	-0.00324*	-0.003	-0.00311**
	(0.0019)	(0.0020)	(0.0014)
Dummy for School District	0.222***	0.206***	0.213***
	(0.0215)	(0.0206)	(0.0152)
Number of Buses	0.00798***	0.00639***	0.00699***
	(0.0017)	(0.0016)	(0.0012)
Dummy for Transfer Station	0.0511**	0.0481**	0.0516***
	(0.0229)	(0.0228)	(0.0165)
Constant	10.53***	10.43***	10.50***
	(0.0592)	(0.0589)	(0.0427)
Observations	731	548	1279
R-squared	0.545	0.663	0.584

Notes: Standard errors in parentheses ***p<0.01,**p<0.05,*p<0.1

Conclusion and Policy Implications

Our regression analysis results show that the completion of the new subway Line 6 has a very significant pulling effect on

the communities along its route. In addition, the pulling effect is approximately the same within 1.5 km, no matter whether the community is located within walking distance (0.8 km) or not, which is contrary to the clustering effect and implies the subway station's high spillover effect. Lastly, the completion of a subway line has the most significant impact on the housing prices in the semi-urban areas.

With regard to policy implications, we summarize some previous studies below for reference.

On planning and financing urban transit investment, Ma, Ye & Titheridge[21] find that properties near rail transit stations command a price premium of around 5%, with this effect being more pronounced in suburban and low-to-middle-income areas, where premiums can reach up to 10%. The study concludes that rail transit investment can effectively reshape urban spatial structures and suggests that a "rail + property development" model could be a viable financing solution for rail transit projects in China.

In addition, Yang's findings[22] reveal the significantly positive effect of improved transport accessibility on property prices, with notable variations across different cities and transport modes. The research highlights the importance of integrating environment-friendly transport into urban planning to enhance property values and promote sustainable urban development. It also suggests that value capture schemes can be effective in financing urban infrastructure investments.

On the relationship between urban planning and transportation infrastructure development, Pan & Li's study[23] underscores the role of urban spatial structure in determining property values, suggesting that areas with better connectivity and infrastructure tend to have higher property prices. The results highlight the importance of integrating urban planning and transportation infrastructure development to enhance property values and promote sustainable urban growth.

Yang, Chen, Xu, Zhao, Chau & Hong's study[24] indicates that the impact of rail transit on property prices is not uniform across the city, with some areas experiencing more significant price increases than others. This spatial heterogeneity suggests that urban planning and value capture strategies should consider local context and characteristics to maximize the benefits of rail transit investments. The study underscores the need for tailored approaches to urban development and transit planning.

Similarly, Costa, Sonnenschein & Zheng's study[25] examines the effects of subway expansions on the geographic concentration of consumer amenities in four global cities. It highlights that subway expansions can exacerbate spatial disparities within cities, as some areas may experience significant benefits while others see little to no improvement. It underscores the importance of considering spatial heterogeneity in urban planning and transit development to ensure equitable distribution of opportunities.

Lastly, in terms of technical suggestions, Li & Huang's findings[26] confirm that rail transit accessibility positively impacts housing prices, with a higher premium observed in areas with more mixed land uses. They suggest that transit-oriented development (TOD) strategies can effectively enhance property values by improving transit accessibility and promoting mixed-use development. The study also provides policy recommendations for TOD planning and land value capture near transit stations.

Hu's research,[27] on the other hand, finds that proximity to rail transit stations generally increases housing prices and reduces dwelling sizes, particularly in suburban areas. This suggests that rail transit investments can influence urban spatial structures by encouraging denser development near stations. The findings highlight the importance of government policies in promoting small-sized and affordable housing around suburban stations to maximize the benefits of rail transit investments.

According to the Tiebout model, the level of public goods provision is capitalized into the house prices. With budget constraints, urban planners aim to maximize social welfare, and house price is an effective indicator of the impact of public transit spending. Our empirical findings suggest that the subway network in Beijing could be more decentralized. Instead of too many subway lines crossing the downtown, building more subway lines leading to the semi-urban centers will have the most effective outcome. Our findings have significant implications for urban planning decisions in other fast-expanding metropolitan areas.

Acknowledgment

I would like to extend my sincere gratitude to Helen Cai, seasoned practitioner and MIT graduate, for her kind mentorship and guidance throughout the development of this paper.

References

1. Baldassare M, Night R, Swan S. (1979). Urban service and environment stressor: The impact of the Bay Area Rapid Transit System on residential mobility. ***Journal of Environment and Behavior***, *11*(4), 435-450. https://10.1177/0013916579114001
2. Prior, P. (1980). The Impact of the New Joban Line on Property Values in Tokyo. ***Journal of Urban Economics***, *8*(1), 105-120.
3. Riley D. (2001). ***Taken for a Ride: Taxpayers***, Trains and HM Treasury. Centre for Land Policy Studies.
4. Xiao Y., Webster C., & Orford S. (2014). Identifying house price effects of changes in urban street configuration: An empirical study in Nanjing, China. ***Journal of Urban Studies***, *53*(1), 112-131. https://doi.org/10.1177/0042098014560500
5. Tan Y., Zhao P., Li L. (2024). Beyond Half-Mile Circle: Measuring the Impact of Subway Expansion on Home-Based Travels in Beijing, China, ***Journal of Transport Policy***, *159*(C), 408-418. https://doi.org/10.1016/j.tranpol.2024.10.027
6. Lai X. (2024). An Empirical Study on the Determinants of Housing Prices in Beijing and Model Optimization. ***Science, Technology and Social Development Proceedings Series 2***, 53-59. https://doi.org/10.70088/8f92z263
7. Wang S. (2024). The Externality Impact of Urban Rail Transit on Urban Real Estate Prices in China. ***Highlights in Science, Engineering and Technology CEAT 2024***, *Volume 86*, 88-94. https://doi.org/10.54097/rszd2j84
8. Tiebout CM. (1956). A Pure Theory of Local Expenditures. ***The Journal of Political Economy***, *64*(5), 416-424. https://doi.org/10.1086/257839
9. Muth R. (1969). ***Cities and Housing: The Spatial Pattern of Urban Residential Land Use***. The University of Chicago Press.

DOI: 10.36838/v7i10.45

10. Allen W.B. (1987). The Impact of the New Jersey Transit System on Housing Prices, ***Journal of the American Institute of Planners***, *35*(1), 50-57.
11. So HM, Tse RYC, Ganesan S. (1997). Estimating the influence of transport on house prices: evidence from Hong Kong. ***Journal of Property Valuation and Investment***, *15*(1), 40-47. https://doi.org/10.1108/14635789710163793
12. Li H, Wei Y, Wu Y & Tian G. (2019). Analyzing housing prices in Shanghai with open data: Amenity, accessibility, and urban structure. ***Journal of Cities***, *91*, 165-179. https://doi.org/10.1016/j.cities.2018.11.016
13. Stegman MA. (1978). The Impact of the Bay Area Rapid Transit System on Residential Property Values, ***Journal of the American Institute of Planners***, *44*(4), 387-396.
14. Sivitanidou R. (1996). Do office-commercial firms value access to service employment centers? A hedonic value analysis within polycentric Los Angeles. ***Journal of Urban Economics***, *40*(2), 125-149.
15. DeFusco A, Nathanson C, & Zwick E. (2022). Speculative Dynamics of Prices and Volume. ***Journal of Financial Economics***, *146*, 205-229. https://doi.org/10.1016/j.jfineco.2022.07.002
16. Rosen S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. ***The Journal of Political Economy***, *82*(1), 34-55. https://doi.org/10.1086/260169
17. Butler, R.V. (1982). The specification of hedonic indexes for urban housing. ***Land Economics***, *58*(1), 96-108. https://doi.org/10.2307/3146079
18. Heenan GW. (1968). The economic effect of rapid transit on real estate development. ***Appraisal Journal***, *36*(2), 213-224.
19. Cockerill L, Stanley D. (2002). ***How Will the Centerline Affect Property Values in Orange County?*** Fullerton: California State University at Fullerton, Institute of Economic and Environmental Studies.
20. White H., A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica: ***Journal of the Econometric Society***, *1980*: 817-838.
21. Ma, L., Ye, R., Titheridge, H. (2014). Capitalization effects of rail transit and bus rapid transit on residential property values in a booming economy: evidence from Beijing. ***Transportation Research Record***, *2451*(1), 139-148. https://doi.org/10.3141/2451-16
22. Yang, L. (2021). ***Property Price Impacts of Environment-Friendly Transport Accessibility in Chinese Cities***. Springer, 9-25. ISBN: 978-981-16-8833-1. https://link.springer.com/book/10.1007/978-981-16-8833-1
23. Pan, H., Li, W. (2019). The impact of urban spatial structure and rail transit on residential property price in Shanghai. ***Transportation Research Procedia***, *41*, 229-230. https://doi.org/10.1016/j.trpro.2019.09.041
24. Yang, L., Chen, Y., Xu, N., Zhao, R., Chau, K.W., Hong, S. (2020). Place-varying impacts of urban rail transit on property prices in Shenzhen, China: Insights for value capture. ***Sustainable Cities and Society***, *58*, 102140. https://doi.org/10.1016/j.scs.2020.102140
25. Costa, A.B., Sonnenschein, T., Zheng, S. (2021). ***Subway expansion and the rise in the spatial disparity of consumer amenities***. SSRN Journal. https://doi.org/10.2139/ssrn.3881397
26. Li, J., Huang, H. (2020). Effects of transit-oriented development (TOD) on housing prices: A case study in Wuhan, China. ***Research in Transportation Economics***, *80*, 100813. https://doi.org/10.1016/j.retrec.2020.100813
27. Hu, R. (2017). The impact of rail transit on the distribution of new housing projects in Beijing. ***Urban Studies***, *54*, 1867-1886. https://doi.org/10.1177/0042098016633085

Author

The author is a senior at the Chinese International School in Hong Kong with a strong interest in economics, urban planning, and policy formation. Eric is also an advocate for environmental sustainability and STEM education, leading initiatives that foster innovation.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Opioid Receptors and Opioid Use Disorder: Implications for Treatment

Esther Cui

Sentinel Secondary, 1250 Chartwell Dr, West Vancouver, BC V7S 2R2, Canada; estherrrcui@gmail.com

ABSTRACT: The global opioid crisis has raised the critical need for effective, safe treatment for opioid use disorder (OUD) with minimal side effects. Central to this research are opioid receptors and signaling mechanisms that influence pain relief, reward, and addiction. This review examines the connection between opioid receptors and OUD, focusing on the mu-opioid receptor (MOR), the delta-opioid receptor (DOR), and the kappa-opioid receptor (KOR). This review also evaluates current medication treatments for OUD (methadone, buprenorphine, naltrexone), highlighting their successes and challenges. Exploring how opioid receptors and their functions contribute to OUD and treatment options can provide insight into mitigating the opioid crisis, guiding the development of safer and more effective OUD therapies.

KEYWORDS: Cellular and Molecular Biology, Neurobiology, Opioid Receptor, Opioid Use Disorder (OUD), Endogenous Opioid System (EOS), Medication-Assisted Treatment (MAT).

Introduction

Opioids are a class of natural or synthetic drugs that derive from or mimic alkaloids found in opium poppies (*Papaver somniferum*). The medicinal and recreational use of naturally occurring opiates for pain relief or to cause euphoria dates back thousands of years. However, it was only in the early 1800s that morphine was isolated as the most active component of opium, making it the first alkaloid to be isolated from a natural product.[1,2] Since then, many more opioids have been isolated and developed to be used clinically as treatment for moderate to severe pain, often prescribed after surgery, injury, or for health conditions like cancer.[3] Since the 1990s, opioid prescription rates have risen significantly. This surge stemmed from the recognition of pain undertreatment as a major clinical issue and the misconception that patients were not at risk of developing opioid use disorder (OUD).[4,5] Despite their side effects, including dependence, euphoria, addiction, respiratory depression, constipation, nausea, and vomiting,[6] opioid use persists today, in part due to the critical need for effective pain management. With estimates suggesting that pain affects 20% of adults globally,[7] opioids are a primary treatment option for acute and cancer pain.

In recent decades, a global public health crisis has emerged, known as the "opioid epidemic" or "opioid crisis." Leading factors of this crisis include the role of pharmaceutical companies, poor regulation, overprescription, and the rise in illegal drug use. A primary reason for the widespread use and misuse of opioids is their high addictive potential. For example, heroin, an illegal opioid made from morphine, is one of the most addictive drugs on the planet.[8] Opioids are so addictive because they interact with the brain's reward system, triggering euphoria and the release of endorphins and dopamine.[9] Consequently, 60 million people worldwide face the addictive effects of opioids,[8] and the crisis is particularly serious in North America. In 2023, pharmacies in the United States dispensed around 125 million opioid prescriptions.[3] Furthermore, in the last two decades, the United States and Canada have had nearly 600,000 deaths caused by opioid overdose. By 2029, this number is estimated to reach 1.2 million.[10] A significant portion of these numbers are linked to OUD, with over 16 million affected by it globally and 2.1 million in the United States.[11]

Characterized by chronic compulsive or harmful use of opioids, OUD is driven by the effects of drugs on the brain, mediated by mu (MOR), kappa (KOR), and delta (DOR) opioid receptors. These receptors were discovered in the early 1970s as the binding site of opioids.[12] They belong to the G protein-coupled receptors (GPCRs) family, the largest class of membrane proteins in the human genome, and some of the most common drug targets. When activated by opioids, these receptors can affect pain perception, mood, and stress, leading to widespread clinical and recreational use.[13] MORs are the most researched out of the three main receptors, as they have been shown to be responsible for the analgesic, rewarding, tolerance-inducing, and withdrawal effects of morphine, a mu receptor agonist. This was demonstrated by a study involving the deletion of OPRM1, the gene encoding MORs, in mice, which subsequently eliminated these effects.[14]

Current pharmacological treatments for OUD focus on reducing withdrawal symptoms, cravings, and the risk of relapse while promoting long-term recovery. So far, three medications have been approved by the FDA for OUD treatment: methadone, buprenorphine, and naltrexone.[15,16] Methadone is a long-acting full MOR agonist that activates the receptor similarly to illicit drugs, but more gradually. Buprenorphine is a long-acting partial MOR agonist, meaning it binds to receptors like full agonists, but with less efficacy. It also displaces other full agonists from binding to the receptor.[17] These two medications are based on a treatment approach known as opioid substitution therapy (OST).[18] When taken correctly, they can reduce withdrawal symptoms and cravings without pro-

DOI: 10.36838/v7i10.53

ducing euphoria. On the other hand, naltrexone is an MOR antagonist, which inhibits the activation of receptors and their effects. It has been used to prolong sobriety, blocking receptors and the euphoric effects of opioids and lowering the chance of relapse.[19,20] However, these current treatments have notable limitations. For example, the high potency and efficacy of methadone increase the risk of overdose, and being a MOR agonist, it may be misused. More importantly, the treatments are accompanied by significant adverse effects, including nausea, vomiting, constipation, and potentially fatal respiratory depression and QT prolongation.[15] These limitations highlight the need for opioid treatments with fewer side effects and greater efficacy in managing OUD.

This review discusses the neurobiological basis in which opioid receptors and their signaling mechanisms contribute to OUD, focusing on their role in reward systems, tolerance, dependence, and other mechanisms underlying addiction. It will explore the benefits and limitations of current treatments of OUD, particularly medication-assisted treatment (MAT). By examining these topics, this review aims to provide a deeper understanding of how advancements in opioid research can address the ongoing opioid crisis.

The Endogenous Opioid System and OUD:

The endogenous opioid system (EOS) is widely distributed throughout the central nervous system (CNS) and peripheral nervous system (PNS), particularly in neural circuits related to pain, reward, emotion, and autonomic control.[21] It comprises three families of receptors (mu, kappa, delta) and opioid peptides acting at these receptors. There are >20 different identified opioid peptides, all of which are processed from three protein precursors: proopiomelanocortin (POMC), prodynorphin (PDYN), and proenkephalin (PENK).[22] All opioid peptides share a common NH2-terminal Tyr-Gly-Gly-Phe signature tetrapeptide sequence known as the "opioid motif",[23] which interacts with the receptors.[24] However, only three main peptides are generally considered to be part of the EOS: β-endorphins, dynorphins, and enkephalins produced through proteolytic cleavage of POMC, PDYN, and PENK, respectively.[24,25] Each peptide binds to all three receptors, albeit while exhibiting different affinities. β-endorphins act primarily on MORs, while dynorphins act on KORs, and enkephalins act on DORs.[26–28]

Opioid Receptor Activation and G-Protein Signaling:

As GPCRs, the three main opioid receptors share a seven-transmembrane (7TM) helical structure in the form of intracellular and extracellular loops.[29] Receptors contain ligand-binding pockets, where ligands—such as endogenous peptides or synthetic drugs—bind to and activate the receptor (see Figure 1). Ligand binding induces conformational changes that enable intracellular coupling of heterotrimeric Gi/o proteins to the receptor's C terminus.[30] The receptor then promotes the exchange of GDP for GTP on the α subunit of the G protein, and the trimeric G protein complex dissociates into Gα and Gβγ. The Gα component inhibits adenylyl cyclase activation, lowering cAMP production, while the Gβγ component interacts with various ion channels.[31,32] Calcium channels are closed, decreasing Ca^{2+} influx, and G protein-coupled inwardly rectifying K+ (GIRK) channels are opened, increasing K+ efflux. These combined actions result in analgesia by causing hyperpolarization and reduced neuronal excitability, as well as diminished nociceptive stimuli and pain perception. Opioid receptor cellular signaling generally occurs through intracellular G proteins and GPCR kinases (GRKs) and β-arrestins.[33] GRKs can phosphorylate an active receptor's C-terminal tail or intracellular loops, promoting β-arrestin recruitment. This leads to receptor desensitization and internalization, which regulates receptor signaling and prevents excessive cellular responses.[1] Repeated opioid use and receptor activation can lead to receptor down-regulation, which, when paired with desensitization and internalization, can contribute to tolerance.

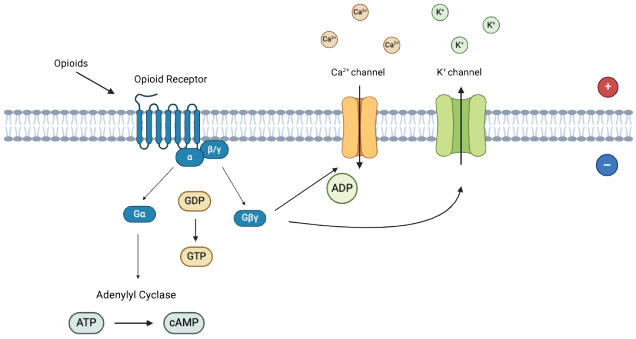


Figure 1: Signal transduction pathway of Gi/o-coupled opioid receptors. The figure illustrates how opioid receptor signaling forms the molecular basis for opioid-induced analgesia and many of their central effects. Upon ligand binding, the receptor undergoes a conformational change that activates the associated Gi/o heterotrimeric G protein. This activation promotes GDP–GTP exchange on the Gα subunit, leading to its dissociation from the Gβγ dimer. The Gα subunit inhibits adenylyl cyclase, reducing cAMP production and downstream intracellular signaling. Simultaneously, the Gβγ subunit modulates ion channel activity by inhibiting voltage-gated calcium (Ca^{2+}) channels and activating G protein-gated inwardly rectifying potassium (GIRK) channels. Created in BioRender. Cui, E. (2025) https://BioRender.com/4d7fnl1

Differential Receptor Function:

Opioid receptors can mediate the effects of endogenous ligands, such as endorphins, as well as exogenous ligands, including morphine and fentanyl.[34] As discussed earlier, these receptors are classified into three main types: the mu-opioid receptor (MOR), kappa-opioid receptor (KOR), and delta-opioid receptor (DOR), each encoded by distinct genes (OPRM1, OPRK1, OPRD1, respectively).[13] While MORs, KORs, and DORs are all implicated in analgesia, studies have shown that they each contribute to further distinct physiological and behavioral effects. In OPRM1 knockout (KO) mice (mice in which the OPRM1 gene has been removed), the effects of morphine on analgesia, reward, withdrawal, dependence development, and respiratory depression were undetected.[14,35] Meanwhile, in OPRK1 KO and OPRD1 KO mice, these effects remained detectable.[36,37] Due to these findings placing the MOR as the responsible receptor for both the therapeutic and the adverse effects of morphine, MORs are the main target for opioid analgesics. On the other hand, DORs have been shown to contribute to mood-related, anxiolytic, and

antidepressant effects,[38] and KORs have been associated with dysphoria, as well as aversive and psychotomimetic effects.[39,40]

Anatomic Distribution of the EOS:

The wide anatomic distribution of the EOS contributes to the analgesic and other physiological effects of opioids on the human body. For example, opioid receptors can be expressed in the lungs, heart, kidneys, pancreas, and small intestine, as well as in neuroendocrine, immune, and ectodermal cells.[41] As a result, they can affect organ function, inflammation, and homeostasis.[42] Opioid receptors are also concentrated in areas of the brain and spinal cord (periaqueductal grey (PAG), locus coeruleus (LC), rostral ventral medulla (RVM), and the spinal dorsal horn (SDH)) that help process and manage pain.[34] In the midbrain, when an opioid binds to the MOR, it triggers signals that inhibit certain neurons, leading to a reduction in pain transmission from the body to the brain. This effect is part of the pain relief system that works through the PAG, which communicates with the SDH, controlling pain signals as they reach the thalamus. In addition, opioid receptors in the dorsal root ganglion (DRG) can help modulate the initial pain signals and suppress DRG activity to reduce pain perception.

Neuroadaptations and OUD:

The EOS is crucial in regulating behaviors related to important survival mechanisms like reward and pain aversion. Thus, when the system's function is compromised, these behaviors are as well. For example, one of the main consequences of such impairment is the risk of developing OUD. Continuous drug use can lead to neuroadaptations, which are changes in the brain's structure and function. This includes the process in which the main initial cellular responses of the brain to a drug adapt to diminish the drug's effects.[43] This means the brain requires more frequent or higher doses of the drug to achieve the same effects, leading to tolerance. Neuroadaptations also include individuals becoming used to the presence of the drug, meaning that in its absence, users will experience withdrawal symptoms like pain, anxiety, and cravings, which encourage continued use.[44] Long-term use can also lead to neuroadaptations in areas of the brain involved in motivation, reward processing, habit formation, and motor control, such as the ventral tegmental area (VTA) and the striatum. An example of a brain alteration caused by a drug of abuse is the expression of stable forms of the ΔFosB protein. In a study done with mice, repeated substance exposure resulted in accumulating levels of ΔFosB in the nucleus accumbens, leading to increased sensitivity to the rewarding effects of the drug. This study suggests that ΔFosB may play a role in developing and maintaining addiction.[45] To conclude, the dysfunction of the EOS in conditions like OUD is linked to its role in modulating the reward-and pain-related effects of substances. These adaptations, occurring in key reward and motivation centers, are critical in the transition from controlled use to addiction and eventually full-blown OUD.

■ Current Pharmacological Treatments for Opioid Use Disorder:

OUD is defined as chronic opioid use despite significant harm or distress and is characterized by neuroadaptation and changes in neuronal circuits.[11] It involves dependence, in which the body adapts to the presence of opioids, which can manifest through withdrawal symptoms (e.g., cravings, sweating, anxiety), prompting one to continue taking opioids. It also involves tolerance, requiring higher or more frequent doses to maintain effects, whether analgesic or euphoric.[46] Treatments of OUD include rehabilitation, cognitive behavioral therapy, and medication-assisted treatment (MAT), which have proved to be particularly effective.[47] Currently, three medications have been approved by the FDA for OUD treatment: methadone, buprenorphine, and naltrexone.[15,16] The first two medications are based on OST and work by replacing the problematic opioid with a safer one.

Table 1: Overview of FDA-approved medications to treat opioid use disorder. The table includes the mechanism of action, effectiveness, benefits, adverse effects, and clinical considerations of methadone, buprenorphine, and naltrexone. While each drug has distinct clinical profiles, methadone and buprenorphine are more strongly associated with reduced mortality and higher treatment retention, whereas naltrexone may be useful for abstinence prolongation but is limited by poor adherence and initiation challenges.[48–57]

Medication	Mechanism of Action	Effectiveness	Benefits	Adverse Effects	Clinical Considerations
Methadone	Full mu-opioid agonist	Associated with decreases in all-cause mortality Higher treatment retention rates than low-dose buprenorphine, comparable to high-dose buprenorphine Effective at reducing opioid-associated transmission of infectious disease,and crime Long-term results (6+ months) more favorable for individuals receiving methadone	Long-acting Reduces withdrawal symptoms and cravings Does not produce euphoria	QTc interval prolongation Risk of respiratory depression Risk of sedation Nausea Constipation Diaphoresis	Patients should be seen daily prior to and during initiation Initial dose: 20-30 mg orally for first dose. In patients with low opioid tolerance, start with 10-15 mg. May give additional 5-10 mg if after 2-4 hours symptoms have not been suppressed or reappear Daily dose: Increase the dose by 10 mg every 5 days as needed, no more than 20 mg a week, to address cravings or withdrawal symptoms More effective at higher daily dosages (80 to 120 mg) than at moderate dosages (40 to 50 mg) Can be administered daily for most patients
Buprenorphine	Partial mu-opioid agonist	All-cause mortality reduced by %50 Must be given at a sufficiently high dose (roughly>16 mg/day) to be effective Patients on doses of buprenorphine of >16 mg/day 1.82 times more likely to stay in treatment than placebo-treated patients	Displays ceiling effect for respiratory depression, euphoria, sedation, and intoxication Reduces cravings Mitigates withdrawal symptoms	May precipitate withdrawal Low risk of respiratory depression Nausea Headache Insomnia Vomiting Constipation	Patients must be in mild to moderate withdrawal before initiating treatment Initial dose: 4-8 mg, depending on patient requirement Daily dose: 4-24 mg, should not exceed 32 mg Overtime, dose should be titrated up to eliminate withdrawal symptoms and reduce cravings Higher doses associated with improved treatment retention and reduced illicit opioid use May initially need to be seen 1-2 times a week, and monthly further on
Naltrexone	Antagonist	Has not been associated with decreased mortality Limited by poor treatment adherence Monthly injections improve adherence; clinical trial showed 90% abstinent weeks vs. 35% with placebo, higher retention (58% vs. 42%) Treatment initiation harder due to required detoxification Extended release naltrexone formulation similar in effectiveness at treating OUD to buprenorphine/naloxone combination	Blocks effect of opioid agonists Does not produce tolerance or withdrawal Not addictive Not at risk of abuse or diversion Suppresses cravings	Headache Lowered tolerance (therefore, increased risk of overdose) Injection site reactions Insomnia Increased alanine transaminase Increased creatine phosphokinase	Patient must be completely detoxicated for 7-14 days before taking naltrexone Initial dose: 25 mg Daily dose: 50 mg Intramuscular naltrexone is administered in doses of 380 mg every four weeks

DOI: 10.36838/v7i10.53

Methadone:

Methadone works as a full MOR agonist and manages withdrawal symptoms (e.g., tachycardia, sweating, nausea, vomiting, and diarrhea), reduces cravings, and dulls the effects of other opioids.[58,59] At a maintenance dose, it does not produce euphoria. Compared to morphine, methadone produces enhanced analgesia with reduced tolerance.[60] This may be due to its inhibition of serotonin and norepinephrine uptake, as well as its role as an antagonist at the N-methyl-D-aspartate (NMDA) receptor, which is involved in pain modulation and transmission.[61] Methadone also has a longer duration of action and half-life, making it a good substitute for pain treatment and for OUD, as fewer doses are required to maintain analgesia and prevent withdrawal symptoms.[62] The effects of methadone will typically last 24-36 hours, while short-acting opioids taken to manage withdrawal will often need to be taken three to four times a day.[55]

Buprenorphine:

Buprenorphine is a partial MOR agonist, meaning the maximal effect buprenorphine can produce will always be lower than that of a full MOR agonist and is thereby less potent. When taken at the proper dosage, buprenorphine can mitigate withdrawal symptoms and reduce cravings without causing euphoria.[63] Buprenorphine has a high affinity for the MOR but lower intrinsic activity compared to full MOR agonists such as morphine, heroin, oxycodone, and methadone. This means that buprenorphine preferentially binds to and displaces other full agonists with lower affinity from the receptor.[17,64] Additionally, buprenorphine has a very low receptor dissociation rate, as well as a long duration of action and half-life.[18] Buprenorphine also appears to display a "ceiling effect" for respiratory depression, euphoria, sedation, and intoxication. This can be beneficial as these effects will plateau, leading to lower risks and less severe effects of overdose compared to full agonists.[19,65]

Naltrexone:

Naltrexone is an antagonist preferentially of the MOR and, to a lesser extent, of the KOR and DOR.[66] Naltrexone will block the euphoric and sedative effects of opioids. As an antagonist, naltrexone is not addictive and has no abuse or diversion potential. Naltrexone is also reported to reduce and suppress cravings and may also reduce tolerance. However, this lowered tolerance could be dangerous in the case of relapse, as taking previous levels of doses could lead to overdose and serious effects like respiratory arrest and circulatory collapse.[67] To reduce the risk of withdrawal caused by OUD, patients must be opioid-free for at least 7-14 days before starting naltrexone.[68] Individuals taking naltrexone should also refrain from using any other opioids or drugs, consuming alcohol, or taking sedatives or tranquilizers. In this way, naltrexone is primarily used to prolong abstinence and maintain recovery.

Naloxone:

Naloxone is a fast-acting antagonist used to rapidly and temporarily reverse overdose rather than directly treat OUD. Naloxone is generally considered to be safe, as there is no risk for misuse or development of dependence. Naloxone has a high affinity for MORs and, acting as an inverse agonist, removes other opioids bound to the receptor.[69] Naloxone is also available commercially in combination with buprenorphine as Suboxone, which is used for maintenance treatment of OUD, dependence, and addiction.[70]

Overall, these medications have proven to be effective in treating OUD and its symptoms, allowing patients to maintain recovery and supporting social reintegration (see Table 1). Substantial evidence has shown that MAT reduces opioid use and OUD-related symptoms, as well as the risk of infectious disease transmission and drug-associated criminal behavior.[20] OST has been associated with considerably lowered all-cause mortality rates[58] and has even been shown to preserve immune and cognitive functions.[71,72] However, they still come with significant limitations (see Table 1). Many patients are still likely to relapse or quit, particularly with naltrexone, as no euphoric or addictive effects are produced. The use of other drugs while taking naltrexone can be incredibly dangerous and carries an increased risk of fatal overdose.[73] Methadone and buprenorphine, as MOR agonists, still have risks of misuse or diversion.[74] Methadone, as a full agonist with high potency and efficacy, has a substantial risk of overdose and pronounced adverse effects, including nausea, vomiting, respiratory depression, pruritus, sedation, hypotension, hypogonadism, constipation, diaphoresis, and QTc interval prolongation.[15,59] Buprenorphine may precipitate withdrawal if taken too soon after a full agonist (e.g., fentanyl, heroin, prescription opioids) by an individual dependent on opioids.[75] This is due to the displacement of other lower-affinity opioids by buprenorphine. Adverse effects of buprenorphine are similar and include nausea, vomiting, memory loss, dizziness, hypotension, CNS depression, constipation, miosis, QTc interval prolongation, respiratory depression, and insomnia.[17,58] Naltrexone can also cause side effects, including nausea, vomiting, anxiety, constipation, insomnia, loss of appetite, dizziness, injection site reactions, increased alanine transaminase (may indicate liver damage or disease), and increased creatine phosphokinase (may indicate muscle, heart, or brain damage).[76,77]

■ Conclusion

Opioid use disorder is closely connected to the interactions of the EOS, particularly through mu, kappa, and delta opioid receptors and their signaling pathways. This review has explored how the EOS and receptors contribute to OUD, particularly through neuroadaptations and their roles in the brain's reward system. This review has also evaluated current treatments—methadone, buprenorphine, and naltrexone—which, while beneficial, have notable limitations. Addressing the opioid crisis requires advancement in two main directions: finding improved analgesics with high efficacy but minimal addictive potential and side effects, and developing new and better treatments for OUD and those facing effects such as dependency, addiction, and withdrawal. In recent years, new promising approaches to treating OUD have emerged, such as neuromodulation and psychedelics. Neuromodulation treatments such as transcranial magnetic stimulation (TMS)

and deep brain stimulation (DBS) target and modulate neural circuits and synaptic plasticity.[78] A systematic review and meta-analysis of neuromodulation therapies for substance use disorders found that TMS produced medium to large effect sizes in reducing substance use and cravings, particularly when multiple stimulation sessions were applied.[79] Additionally, psychedelic medicine, such as psilocybin and LSD, has reemerged as a topic of discussion and possible therapy. They are generally considered to have low potential for dependence and addiction, and several studies have produced results suggesting psychedelic use to be associated with reduced odds of problematic substance use or OUD.[80–83] Other approaches include targeting components of the dopamine-dependent reward circuitry, identifying genetic factors that increase vulnerability to OUD, and modulating gene products accordingly. Vaccines are also being explored as a form of immunotherapy by reducing or slowing drug entry into the brain, thereby reducing effects associated with overdose.[84–87] While further research and trials are necessary to ensure the safety and effectiveness of these treatments, they show promise toward offering new methods to manage opioid use disorder and the larger opioid crisis.

Acknowledgments

I want to thank Dr. Shaye and Dr. Tetz for their incredibly patient guidance throughout my research. I'm very grateful for the support they've provided me.

References

1. Che, T.; Roth, B. L. Molecular Basis of Opioid Receptor Signaling. *Cell* **2023**, *186* (24), 5203–5219. https://doi.org/10.1016/j.cell.2023.10.029.
2. Krishnamurti, C.; Rao, S. C. The Isolation of Morphine by Serturner. *Indian J. Anaesth.* **2016**, *60* (11), 861. https://doi.org/10.4103/0019-5049.193696.
3. *About Prescription Opioids | Overdose Prevention | CDC*. https://www.cdc.gov/overdose-prevention/about/prescription-opioids.html.
4. Addiction Rare in Patients Treated with Narcotics. *N. Engl. J. Med.* **1980**, *302* (2), 123–123. https://doi.org/10.1056/NEJM198001103020221.
5. Institute of Medicine (US) Committee on Advancing Pain Research, Care, and Education. *Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research*; The National Academies Collection: Reports funded by National Institutes of Health; National Academies Press (US): Washington (DC), 2011.
6. Machelska, H.; Celik, M. Ö. Advances in Achieving Opioid Analgesia Without Side Effects. *Front. Pharmacol.* **2018**, *9*. https://doi.org/10.3389/fphar.2018.01388.
7. Goldberg, D. S.; McGee, S. J. Pain as a Global Public Health Priority. *BMC Public Health* **2011**, *11* (1), 770. https://doi.org/10.1186/1471-2458-11-770.
8. Americas, T. L. R. H.-. Opioid Crisis: Addiction, Overprescription, and Insufficient Primary Prevention. *Lancet Reg. Health – Am.* **2023**, *23*. https://doi.org/10.1016/j.lana.2023.100557.
9. Kosten, T. R.; George, T. P. The Neurobiology of Opioid Dependence: Implications for Treatment. *Sci. Pract. Perspect.* **2002**, *1* (1), 13–20. https://doi.org/10.1151/spp021113.
10. Health, T. L. P. Opioid Overdose Crisis: Time for a Radical Rethink. *Lancet Public Health* **2022**, *7* (3), e195. https://doi.org/10.1016/S2468-2667(22)00043-3.
11. Dydyk, A. M.; Jain, N. K.; Gupta, M. Opioid Use Disorder. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2025.
12. Fairbanks, C. A.; Peterson, C. D. The Opioid Receptor: Emergence through Millennia of Pharmaceutical Sciences. *Front. Pain Res.* **2023**, *4*. https://doi.org/10.3389/fpain.2023.960389.
13. Darcq, E.; Kieffer, B. L. Opioid Receptors: Drivers to Addiction? *Nat. Rev. Neurosci.* **2018**, *19* (8), 499–514. https://doi.org/10.1038/s41583-018-0028-x.
14. Matthes, H. W. D.; Maldonado, R.; Simonin, F.; Valverde, O.; Slowe, S.; Kitchen, I.; Befort, K.; Dierich, A.; Le Meur, M.; Dollé, P.; Tzavara, E.; Hanoune, J.; Roques, B. P.; Kieffer, B. L. Loss of Morphine-Induced Analgesia, Reward Effect and Withdrawal Symptoms in Mice Lacking the μ-Opioid-Receptor Gene. *Nature* **1996**, *383* (6603), 819–823. https://doi.org/10.1038/383819a0.
15. Lu, T.; Whitley, S. D.; Wiegand, T. J.; Stancliff, S. L.; Norton, B. L.; Hoffmann, C. J.; Gonzalez, C. J. *Treatment of Opioid Use Disorder*; New York State Department of Health AIDS Institute Clinical Guidelines; Johns Hopkins University: Baltimore (MD), 2024.
16. Research, C. for D. E. and. Information about Medications for Opioid Use Disorder (MOUD). *FDA* **2024**.
17. Kumar, R.; Viswanath, O.; Saadabadi, A. Buprenorphine. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2025.
18. Noble, F.; Marie, N. Management of Opioid Addiction With Opioid Substitution Treatments: Beyond Methadone and Buprenorphine. *Front. Psychiatry* **2019**, *9*, 742. https://doi.org/10.3389/fpsyt.2018.00742.
19. Bell, J.; Strang, J. Medication Treatment of Opioid Use Disorder. *Biol. Psychiatry* **2020**, *87* (1), 82–88. https://doi.org/10.1016/j.biopsych.2019.06.020.
20. Abuse, N. I. on D. How do medications to treat opioid use disorder work? | *National Institute on Drug Abuse (NIDA)*. https://nida.nih.gov/publications/research-reports/medications-to-treat-opioid-addiction/how-do-medications-to-treat-opioid-addiction-work.
21. Benarroch, E. E. Endogenous Opioid Systems: Current Concepts and Clinical Correlations. *Neurology* **2012**, *79* (8), 807–814. https://doi.org/10.1212/WNL.0b013e3182662098.
22. Gomes, I.; Sierra, S.; Lueptow, L.; Gupta, A.; Gouty, S.; Margolis, E. B.; Cox, B. M.; Devi, L. A. Biased Signaling by Endogenous Opioid Peptides. *Proc. Natl. Acad. Sci.* **2020**, *117* (21), 11820–11828. https://doi.org/10.1073/pnas.2000712117.
23. Wang, Y.; Zhuang, Y.; DiBerto, J. F.; Zhou, X. E.; Schmitz, G. P.; Yuan, Q.; Jain, M. K.; Liu, W.; Melcher, K.; Jiang, Y.; Roth, B. L.; Xu, H. E. Structures of the Entire Human Opioid Receptor Family. *Cell* **2023**, *186* (2), 413-427.e17. https://doi.org/10.1016/j.cell.2022.12.026.
24. Le Merrer, J.; Becker, J. A. J.; Befort, K.; Kieffer, B. L. Reward Processing by the Opioid System in the Brain. *Physiol. Rev.* **2009**, *89* (4), 1379–1412. https://doi.org/10.1152/physrev.00005.2009.
25. Peciña, M.; Karp, J. F.; Mathew, S.; Todtenkopf, M. S.; Ehrich, E. W.; Zubieta, J.-K. Endogenous Opioid System Dysregulation in Depression: Implications for New Therapeutic Approaches. *Mol. Psychiatry* **2019**, *24* (4), 576–587. https://doi.org/10.1038/s41380-018-0117-2.
26. Cullen, J. M.; Cascella, M. Physiology, Enkephalin. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2025.
27. Pilozzi, A.; Carro, C.; Huang, X. Roles of β-Endorphin in Stress, Behavior, Neuroinflammation, and Brain Energy Metabolism. *Int. J. Mol. Sci.* **2021**, *22* (1), 338. https://doi.org/10.3390/ijms22010338.

DOI: 10.36838/v7i10.53

28. Schwarzer, C. 30 Years of Dynorphins — New Insights on Their Functions in Neuropsychiatric Diseases. *Pharmacol. Ther.* **2009**, *123* (3), 353–370. https://doi.org/10.1016/j.pharmthera.2009.05.006.
29. Serohijos, A. W. R.; Yin, S.; Ding, F.; Gauthier, J.; Gibson, D. G.; Maixner, W.; Dokholyan, N. V.; Diatchenko, L. Structural Basis for μ-Opioid Receptor Binding and Activation. *Structure* **2011**, *19* (11), 1683–1690. https://doi.org/10.1016/j.str.2011.08.003.
30. Stein, C. Opioid Receptors. *Annu. Rev. Med.* **2016**, *67* (Volume 67, 2016), 433–451. https://doi.org/10.1146/annurev-med-062613-093100.
31. Lüscher, C.; Slesinger, P. A. Emerging Roles for G Protein-Gated Inwardly Rectifying Potassium (GIRK) Channels in Health and Disease. *Nat. Rev. Neurosci.* **2010**, *11* (5), 301–315. https://doi.org/10.1038/nrn2834.
32. Tedford, H. W.; Zamponi, G. W. Direct G Protein Modulation of Cav2 Calcium Channels. *Pharmacol. Rev.* **2006**, *58* (4), 837–862. https://doi.org/10.1124/pr.58.4.11.
33. Wacker, D.; Stevens, R. C.; Roth, B. L. How Ligands Illuminate GPCR Molecular Pharmacology. *Cell* **2017**, *170* (3), 414–427. https://doi.org/10.1016/j.cell.2017.07.009.
34. Dhaliwal, A.; Gupta, M. Physiology, Opioid Receptor. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2025.
35. Kieffer, B. L.; Kieffer, B. L. Opioids: First Lessons from Knockout Mice. *Trends Pharmacol. Sci.* **1999**, *20* (1), 19–26. https://doi.org/10.1016/S0165-6147(98)01279-6.
36. Simonin, F.; Valverde, O.; Smadja, C.; Slowe, S.; Kitchen, I.; Dierich, A.; Le Meur, M.; Roques, B. P.; Maldonado, R.; Kieffer, B. L. Disruption of the K-opioid Receptor Gene in Mice Enhances Sensitivity to Chemical Visceral Pain, Impairs Pharmacological Actions of the Selective K-agonist U-50,488H and Attenuates Morphine Withdrawal. *EMBO J.* **1998**, *17* (4), 886–897. https://doi.org/10.1093/emboj/17.4.886.
37. Zhu, Y.; King, M. A.; Schuller, A. G. P.; Nitsche, J. F.; Reidl, M.; Elde, R. P.; Unterwald, E.; Pasternak, G. W.; Pintar, J. E. Retention of Supraspinal Delta-like Analgesia and Loss of Morphine Tolerance in δ Opioid Receptor Knockout Mice. *Neuron* **1999**, *24* (1), 243–252. https://doi.org/10.1016/S0896-6273(00)80836-3.
38. Filliol, D.; Ghozland, S.; Chluba, J.; Martin, M.; Matthes, H. W. D.; Simonin, F.; Befort, K.; Gavériaux-Ruff, C.; Dierich, A.; LeMeur, M.; Valverde, O.; Maldonado, R.; Kieffer, B. L. Mice Deficient for δ- and μ-Opioid Receptors Exhibit Opposing Alterations of Emotional Responses. *Nat. Genet.* **2000**, *25* (2), 195–200. https://doi.org/10.1038/76061.
39. Dalefield, M. L.; Scouller, B.; Bibi, R.; Kivell, B. M. The Kappa Opioid Receptor: A Promising Therapeutic Target for Multiple Pathologies. *Front. Pharmacol.* **2022**, *13*. https://doi.org/10.3389/fphar.2022.837671.
40. Pfeiffer, A.; Brantl, V.; Herz, A.; Emrich, H. M. Psychotomimesis Mediated by Kappa Opiate Receptors. Science 1986, 233 (4765), 774–776. https://doi.org/10.1126/science.3016896.
41. Higginbotham, J. A.; Markovic, T.; Massaly, N.; Morón, J. A. Endogenous Opioid Systems Alterations in Pain and Opioid Use Disorder. *Front. Syst. Neurosci.* **2022**, *16*. https://doi.org/10.3389/fnsys.2022.1014768.
42. Peng, J.; Sarkar, S.; Chang, S. L. Opioid Receptor Expression in Human Brain and Peripheral Tissues Using Absolute Quantitative Real-Time RT-PCR. *Drug Alcohol Depend.* **2012**, *124* (3), 223–228. https://doi.org/10.1016/j.drugalcdep.2012.01.013.
43. Wise, R. A.; Koob, G. F. The Development and Maintenance of Drug Addiction. *Neuropsychopharmacology* **2014**, *39* (2), 254–262. https://doi.org/10.1038/npp.2013.261.
44. Koob, G. F.; Volkow, N. D. Neurocircuitry of Addiction. *Neuropsychopharmacology* **2010**, *35* (1), 217–238. https://doi.org/10.1038/npp.2009.110.
45. Kelz, M. B.; Chen, J.; Carlezon, W. A.; Whisler, K.; Gilden, L.; Beckmann, A. M.; Steffen, C.; Zhang, Y.-J.; Marotti, L.; Self, D. W.; Tkatch, T.; Baranauskas, G.; Surmeier, D. J.; Neve, R. L.; Duman, R. S.; Picciotto, M. R.; Nestler, E. J. Expression of the Transcription Factor ΔFosB in the Brain Controls Sensitivity to Cocaine. *Nature* **1999**, *401* (6750), 272–276. https://doi.org/10.1038/45790.
46. Ballantyne, J. C.; Sullivan, M. D.; Kolodny, A. Opioid Dependence vs Addiction: A Distinction Without a Difference? *Arch. Intern. Med.* **2012**, *172* (17), 1342–1343. https://doi.org/10.1001/archinternmed.2012.3212.
47. CDC. *Treatment of Opioid Use Disorder*. Overdose Prevention. https://www.cdc.gov/overdose-prevention/treatment/opioid-use-disorder.html.
48. Mattick, R. P.; Breen, C.; Kimber, J.; Davoli, M. Methadone Maintenance Therapy versus No Opioid Replacement Therapy for Opioid Dependence - Mattick, RP - 2009 | Cochrane Library.
49. Mattick, R. P.; Breen, C.; Kimber, J.; Davoli, M. Buprenorphine Maintenance versus Placebo or Methadone Maintenance for Opioid Dependence - Mattick, RP - 2014 | Cochrane Library.
50. Schwartz, R. P.; Highfield, D. A.; Jaffe, J. H.; Brady, J. V.; Butler, C. B.; Rouse, C. O.; Callaman, J. M.; O'Grady, K. E.; Battjes, R. J. A Randomized Controlled Trial of Interim Methadone Maintenance. *Arch. Gen. Psychiatry* **2006**, *63* (1), 102–109. https://doi.org/10.1001/archpsyc.63.1.102.
51. Vanichseni, S.; Wongsuwan, B.; Choopanya, K.; Wongpanich, K. A Controlled Trial of Methadone Maintenance in a Population of Intravenous Drug Users in Bangkok: Implications for Prevention of HIV. *Int. J. Addict.* **1991**, *26* (12), 1313–1320. https://doi.org/10.3109/10826089109062163.
52. Nunes, E. V.; Krupitsky, E.; Ling, W.; Zummo, J.; Memisoglu, A.; Silverman, B. L.; Gastfriend, D. R. Treating Opioid Dependence With Injectable Extended-Release Naltrexone (XR-NTX): Who Will Respond? *J. Addict. Med.* **2015**, *9* (3), 238. https://doi.org/10.1097/ADM.0000000000000125.
53. Minozzi, S.; Amato, L.; Vecchi, S.; Davoli, M.; Kirchmayer, U.; Verster, A. Oral Naltrexone Maintenance Treatment for Opioid Dependence - Minozzi, S - 2011 | Cochrane Library.
54. Krupitsky, E.; Nunes, E. V.; Ling, W.; Illeperuma, A.; Gastfriend, D. R.; Silverman, B. L. Injectable Extended-Release Naltrexone for Opioid Dependence: A Double-Blind, Placebo-Controlled, Multicentre Randomised Trial. *The Lancet* **2011**, *377* (9776), 1506–1513. https://doi.org/10.1016/S0140-6736(11)60358-9.
55. *Methadone Dosage Guide + Max Dose, Adjustments*. Drugs.com. https://www.drugs.com/dosage/methadone.html.
56. TIP 63: Medications for Opioid Use Disorder.
57. Prescribing Guidelines. In *Guidelines for the Psychosocially Assisted Pharmacological Treatment of Opioid Dependence*, World Health Organization, 2009.
58. Coffa, D.; Snyder, H. Opioid Use Disorder: Medical Treatment Options. *Am. Fam. Physician* **2019**, *100* (7), 416–425.
59. Durrani, M.; Bansal, K. Methadone. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2025.
60. Walwyn, W. M.; Miotto, K. A.; Evans, C. J. Opioid Pharmaceuticals and Addiction: The Issues, and Research Directions Seeking Solutions. *Drug Alcohol Depend.* **2010**, *108* (3), 156–165. https://doi.org/10.1016/j.drugalcdep.2010.01.001.
61. Dedek, A.; Hildebrand, M. E. Advances and Barriers in Understanding Presynaptic N-Methyl-D-Aspartate Receptors in Spinal Pain Processing. *Front. Mol. Neurosci.* **2022**, *15*. https://doi.org/10.3389/fnmol.2022.864502.
62. *Methadone*. https://go.drugbank.com/drugs/DB00333.
63. *Buprenorphine Opioid Addiction Treatment.* CAMH.

https://www.camh.ca/en/health-info/mental-illness-and-addiction-index/buprenorphine.
64. Lutfy, K.; Cowan, A. Buprenorphine: A Unique Drug with Complex Pharmacology. *Curr. Neuropharmacol. 2* (4), 395–402. https://doi.org/10.2174/1570159043359477.
65. Richardson, M. G.; Raymond, B. L. Lack of Evidence for Ceiling Effect for Buprenorphine Analgesia in Humans. *Anesth. Analg.* **2018**, *127* (1), 310. https://doi.org/10.1213/ANE.0000000000003368.
66. Niciu, M. J.; Arias, A. J. Targeted Opioid Receptor Antagonists in the Treatment of Alcohol Use Disorders. *CNS Drugs* **2013**, *27* (10), 777–787. https://doi.org/10.1007/s40263-013-0096-4.
67. *What is Naltrexone? Side Effects, Treatments & Use*. https://www.samhsa.gov/substance-use/treatment/options/naltrexone.
68. Singh, D.; Saadabadi, A. Naltrexone. In *StatPearls*; StatPearls Publishing: Treasure Island (FL), 2025.
69. *Naloxone dosage for opioid reversal: current evidence and clinical implications - Rachael Rzasa Lynn, JL Galinkin, 2018*. https://journals.sagepub.com/doi/10.1177/2042098617744161.
70. *Naloxone*. https://go.drugbank.com/drugs/DB01183.
71. Elkana, O.; Adelson, M.; Doniger, G. M.; Sason, A.; Peles, E. Cognitive Function Is Largely Intact in Methadone Maintenance Treatment Patients. *World J. Biol. Psychiatry* **2019**, *20* (3), 219–229. https://doi.org/10.1080/15622975.2017.1342047.
72. Sacerdote, P.; Franchi, S.; Gerra, G.; Leccese, V.; Panerai, A. E.; Somaini, L. Buprenorphine and Methadone Maintenance Treatment of Heroin Addicts Preserves Immune Function. *Brain. Behav. Immun.* **2008**, *22* (4), 606–613. https://doi.org/10.1016/j.bbi.2007.12.013.
73. Degenhardt, L.; Larney, S.; Kimber, J.; Farrell, M.; Hall, W. Excess Mortality among Opioid-Using Patients Treated with Oral Naltrexone in Australia. *Drug Alcohol Rev.* **2015**, *34* (1), 90–96. https://doi.org/10.1111/dar.12205.
74. Wright, N.; D'Agnone, O.; Krajci, P.; Littlewood, R.; Alho, H.; Reimer, J.; Roncero, C.; Somaini, L.; Maremmani, I. Addressing Misuse and Diversion of Opioid Substitution Medication: Guidance Based on Systematic Evidence Review and Real-World Experience. *J. Public Health* **2016**, *38* (3), e368–e374. https://doi.org/10.1093/pubmed/fdv150.
75. Varshneya, N. B.; Thakrar, A. P.; Hobelmann, J. G.; Dunn, K. E.; Huhn, A. S. Evidence of Buprenorphine-Precipitated Withdrawal in Persons Who Use Fentanyl. *J. Addict. Med.* **2022**, *16* (4), e265. https://doi.org/10.1097/ADM.0000000000000922.
76. Naltrexone. In *LiverTox: Clinical and Research Information on Drug-Induced Liver Injury*; National Institute of Diabetes and Digestive and Kidney Diseases: Bethesda (MD), 2012.
77. *Naltrexone: MedlinePlus Drug Information*. https://medlineplus.gov/druginfo/meds/a685041.html.
78. Lee, Y. K.; Gold, M. S.; Blum, K.; Thanos, P. K.; Hanna, C.; Fuehrlein, B. S. Opioid Use Disorder: Current Trends and Potential Treatments. *Front. Public Health* **2024**, *11*. https://doi.org/10.3389/fpubh.2023.1274719.
79. Mehta, D. D.; Praecht, A.; Ward, H. B.; Sanches, M.; Sorkhou, M.; Tang, V. M.; Steele, V. R.; Hanlon, C. A.; George, T. P. A Systematic Review and Meta-Analysis of Neuromodulation Therapies for Substance Use Disorders. *Neuropsychopharmacology* **2024**, *49* (4), 649–680. https://doi.org/10.1038/s41386-023-01776-0.
80. Argento, E.; Socias, M. E.; Hayashi, K.; Choi, J.; Mackay, L.; Christie, D.; Milloy, M.-J.; DeBeck, K. Psychedelic Use Is Associated with Reduced Daily Opioid Use among People Who Use Illicit Drugs in a Canadian Setting. *Int. J. Drug Policy* **2022**, *100*, 103518. https://doi.org/10.1016/j.drugpo.2021.103518.
81. Garcia-Romeu, A.; Davis, A. K.; Erowid, E.; Erowid, F.; Griffiths, R. R.; Johnson, M. W. Persisting Reductions in Cannabis, Opioid, and Stimulant Misuse After Naturalistic Psychedelic Use: An Online Survey. *Front. Psychiatry* **2020**, *10*. https://doi.org/10.3389/fpsyt.2019.00955.
82. Jones, G.; Ricard, J. A.; Lipson, J.; Nock, M. K. Associations between Classic Psychedelics and Opioid Use Disorder in a Nationally-Representative U.S. Adult Sample. *Sci. Rep.* **2022**, *12* (1), 4099. https://doi.org/10.1038/s41598-022-08085-4.
83. Shen, M. R.; Campbell, D. E.; Kopczynski, A.; Maddams, S.; Rosenblum, N.; Nigam, K.; Suzuki, J. Ketamine in Treating Opioid Use Disorder and Opioid Withdrawal: A Scoping Review. *Front. Psychiatry* **2025**, *16*. https://doi.org/10.3389/fpsyt.2025.1552084.
84. Lee, Y. K.; Gold, M. S.; Fuehrlein, B. S. Looking beyond the Opioid Receptor: A Desperate Need for New Treatments for Opioid Use Disorder. *J. Neurol. Sci.* **2022**, *432*. https://doi.org/10.1016/j.jns.2021.120094.
85. Crist, R. C.; Reiner, B. C.; Berrettini, W. H. A Review of Opioid Addiction Genetics. *Curr. Opin. Psychol.* **2019**, *27*, 31–35. https://doi.org/10.1016/j.copsyc.2018.07.014.
86. Blanco, C.; Volkow, N. D. Management of Opioid Use Disorder in the USA: Present Status and Future Directions. *The Lancet* **2019**, *393* (10182), 1760–1772. https://doi.org/10.1016/S0140-6736(18)33078-2.
87. Pravetoni, M.; Comer, S. D. Development of Vaccines to Treat Opioid Use Disorders and Reduce Incidence of Overdose. *Neuropharmacology* **2019**, *158*, 107662. https://doi.org/10.1016/j.neuropharm.2019.06.001.

Author

Esther Cui is a 10th grader at Sentinel Secondary. In her free time, she can be found crocheting or cooking.

DOI: 10.36838/v7i10.53

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Potential Treatment for Alzheimer's Disease: CRISPR/Cas9

Iván Suárez Pérez

Tecnológico de Monterrey, México; ivansuarez.sp@gmail.com

ABSTRACT: Alzheimer's disease (AD) is a multifactorial neurodegenerative disease caused by a combination of risk factors such as increasing age, genetic factors, vascular diseases, diabetes, infections, and air pollution. AD is a concerning disease as it is the leading cause of dementia and currently lacks a cure. It affects approximately 50 million individuals around the world. Up to 70% of early-onset Alzheimer's disease (EOAD) cases had genetic factors that were related to the disease. Mutations in Amyloid precursor protein (APP), Presenilin-1 (PSEN-1), and Presenilin-2 (PSEN-2) were linked with the development of familial AD. Currently approved drugs are effective only in treating the symptoms of AD, facilitating the synaptic function; they do not cure or prevent the disease. In this paper, we detail the current clinical treatments for AD and biomarkers such as cerebrospinal fluid (CSF) used to detect the formation of beta-amyloid, which are highly related to mutations. This paper aims to evaluate the availability of CRISPR/Cas9 as a treatment for familial Alzheimer's disease (FAD), exploring the technological challenges involved in its implementation. An understanding of potential treatments for AD and their limitations can contribute to the future development of a treatment involving CRISPR/Cas9 that cures AD.

KEYWORDS: Cellular and Molecular Biology, Genetics, Alzheimer's Disease (AD), CRISPR/Cas9, Gene editing, Mutations, Disease Treatment and Therapies.

Introduction

One of the most concerning types of dementia is Alzheimer's disease (AD).[1] AD is a chronic neurodegenerative condition characterized by neuronal death, which affects memory and cognition.[2] Around the world, an estimated 50 million people suffer from AD, and projections show that by 2050, the number of patients with AD will go up to 150 million. The estimated annual cost of AD rises to US$1 trillion, considering the individuals affected, their families, and the economy.[1,3] A major concern is that the causes of AD are multifactorial, with increasing age, vascular diseases, and diabetes being major risk factors for developing the disease. Most AD cases start after 65 years of age; however, in early onset AD (EOAD), genetic factors were related to 70% in the inheritance of it. AD can be divided into early-onset AD (EOAD), which is around 1-6% of cases, with a range of 30-60 years old, where most cases are related to familial AD. Most FAD cases are inherited in an autosomal dominant pattern, with mutations in the dominant genes such as Amyloid precursor protein (APP), Presenilin-1 (PSEN-1), and Presenilin-2 (PSEN-2). Late-onset AD (LOAD) presents more frequently after 65 years old, the multifactorial risks were highly related to AD, and less than 1% of LOAD cases were related to mutations in the apolipoprotein E (ApoE) gene, making gene editing with CRISPR/Cas9 irrelevant. AD lacks a cure, and the current available treatments for it are limited to just improving the symptoms, including side effects such as depression, dizziness, and diarrhea.[3-5] In AD, the neurons die because of structural and functional damage in the central nervous system (CNS), leading to a decrease in the brain size (Figure 1b). The Amyloid beta and Tau hypothesis pathways were suggested to explain the most common hallmarks of AD. In amyloid beta hypothesis structural damage is caused by neuritic plaques that generate due to the accumulation of amyloid-beta peptide's (Aβ) (Figure 2a), which gathers abnormally outside nerve cells; and neurofibrillary tangles (NFT) made up of tau protein, which normally stabilizes the microtubules in healthy neurons, but in AD phosphorylated Tau protein detaches from the microtubules and forms tangles that blocks the neurons communication system and cause their posterior death. (Figure 2b).[1,6] Each current potential treatment for AD faces challenges, but some treatments are more likely to become a cure for AD. A potential treatment that could be used for AD is the early detection of mutations with biomarkers and the subsequent design of a single-guide RNA that recognizes mutated DNA. This will be corrected via clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9 (CRISPR/Cas9). Although there are some limiting factors for this tool, like the safe delivery methods that currently exist. Viral and non-viral methods are used to deliver CRISPR/Cas9 with a lack of efficiency.[3] CRISPR/Cas9 is a revolutionary tool that targets the mutations that relate to AD; it provides a pathway for treating diseases with limited or scarce treatment options, such as AD and Huntington's disease (HD).[3] Currently, CRISPR/Cas9 has not yet been successfully used to cure AD, although it has been used successfully in other diseases, indicating potential for future applications in AD.

In this paper, we first summarize what Alzheimer's disease is, mutations of the presenilin-1 (PSEN1) gene, presenilin 2 (PSEN2) gene, and amyloid precursor protein (APP) that are highly linked with AD. The symptomatic treatments used for AD as cholinesterase inhibitors and N-methyl d-aspartate (NMDA) antagonists. We discuss the potential use of CRISPR/Cas9 and its challenges as brain delivery via non-vi-

DOI: 10.36838/v7i10.60

ral vectors, brain delivery via viral vectors, immunogenicity, blood-brain barrier (BBB), and reticuloendothelial system (RES). This paper aims to discuss whether CRISPR-Cas9 is a viable treatment that could be used for FAD, as this tool has been used to cure other diseases, and current technological challenges could be addressed in the future by research, bringing a way that alleviate the symptoms of FAD.

Discussion

Aβ and Tau protein processing:

Neuritic plaques (NPs) and neurofibrillary tangles (NFTs) are highly involved in several neuronal processes, such as alteration in the permeability of the blood-brain barrier (BBB), mitochondrial disturbance, neuroinflammation, neurovascular unit dysfunction, oxidative stress, and synaptic alteration.[1]

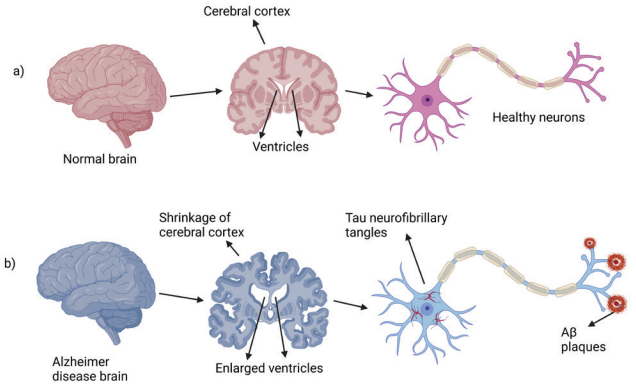


Figure 1: Structure of a healthy brain in (a) and structure of a brain with AD in (b). The AD brain shows significant atrophy, especially in areas responsible for memory and cognition. (Created by author).[2]

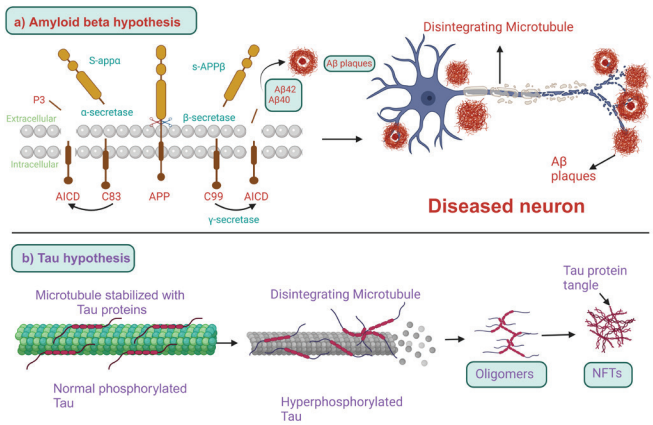


Figure 2: Hypothesis of Aβ relation with AD in (a) and hypothesis of tau protein misfolding related to AD in (b). These figures show how abnormal protein accumulation contributes to neuronal damage and cognitive decline. (Created by author)

NPs and NFT have been postulated to act as endogenous "damage signals" because Aβ oligomers activate microglial cells. Once activated, these cells release proinflammatory cytokines that contribute to neuronal damage. Tau protein is released when neurons die, which can further activate microglial cells, perpetuating a cycle of inflammation, tau hyperphosphorylation, and subsequent aggregation. This self-reinforcing cycle leads to progressive neurodegeneration.[1]

Mutations that cause Alzheimer's disease:

- ***Presenilin-1 (PSEN1) gene:***

A main causative factor of early-onset Alzheimer's disease (EOAD) is the presenilin-1 (PSEN1) gene. The PSEN1 gene codes for the core protein presenilin-1, which is responsible for cutting longer proteins. Presenilin-1 interacts in this way with the amyloid precursor protein (APP), Notch, nicastrin, and the modifier of cellular adhesion (MOCA), beta-catenin. PSEN1 activates the y-secretase complex that plays a role in the production of beta-amyloid from amyloid precursor protein.1 Mutations of PSEN1 were confirmed to impact long amyloid (Aβ42) production and reduce the production of short amyloid (Aβ40); these mutations increase the ratio of Aβ42/Aβ40.[8]

Mutations in the PSEN1 gene provoke the overproduction of long amyloid Aβ42, which tends to aggregate and deposit as plaque in the brain. The cerebrospinal fluid (CSF) is in direct communication with the brain, and when there is an overproduction of Aβ42 and deposition into plaques, the CSF concentration of Aβ42 drops. A low concentration of Aβ42 in the CSF is an indicator of AD.

- ***Presenilin 2 (PSEN2) gene:***

Presenilin 1 and Presenilin 2 share the catalytic core of the y-secretase complex, the enzyme responsible for the generation of beta-amyloid peptides.[9] Presenilin 2, as Presenilin-1 cuts longer proteins. Mutations on PSEN 2 are less common than in PSEN 1. About 300 mutations for PSEN 1 have been described, and only 58 mutations in PSEN 2, which are described as rare and playing a minor role in beta-amyloid production.[1,9] Nevertheless, any mutation in PSEN 2 has severe effects on the Ab42/Ab40 ratio level, increasing the production of Ab42, building up in the brain and forming amyloid plaques as Aβ42 is a protein fragment and is more likely to aggregate than Aβ40, and reducing the concentration of Aβ42 in the CSF.[1-9] Mutations in PSEN 1 and PSEN 2 affect the production of beta-amyloid, leading to the irregular production of long amyloid (Aβ42), which forms amyloid plaques in the brain.

Amyloid precursor protein (APP):

The APP gene is encoded on chromosome 21. It is a type 1 transmembrane protein cleaved by y-secretase to release beta-amyloid and other proteins. Mutations on the APP gene are less common than in PSEN1 and PSEN2; around 25 mutations cause beta-amyloid production and accumulation in the brain.[1] One protective mutation on the APP gene has been identified with the name Icelandic mutation (A673T), decreasing beta-amyloid Ab40 and Ab42 secretion into the brain, reducing the chance of plaque deposition. Another important mutation on APP is A673V, which demonstrates the presence of NFTs and overproduction of beta-amyloid that contribute to neuronal loss.[1] Tables 1, 2, and 3 (see Annex) summarize key mutations in APP, PSEN 1, and PSEN 2 genes, including codon positions and associated clinical relevance for potential gene editing using CRISPR/Cas9.

Current Treatments for Alzheimer's:

The current treatments for AD are reduced to only two classes of drugs, one being inhibitors to cholinesterase enzymes and antagonists to N-methyl-d-aspartate (NMDA).

One of the multiple factors verified to contribute to the development of AD is the reduction in acetylcholine (ACh) biosynthesis. The therapeutic strategy used to address this is by inhibiting cholinesterase enzymes, also known as acetylcholinesterase (AChE), with cholinesterase inhibitors (ChE-Is). The inhibition of AChE stops the degradation in synapses. The increase in ACh results in its continuous accumulation, which activates cholinergic receptors. This treatment was considered to increase the cognitive and neural cells' functions.[1,6] NMDA is the other drug approved to treat AD and is believed to have an important role in the pathophysiology of the disease. NMDA is a membrane receptor that participates in the transmission of electric pulses into the CNS. The stimulation of NMDA results in the formation of long-term potentiation (LTP), which is important for synaptic neurotransmission, plasticity, and memory formation. Verified side effects are caused by the over-activation of NMDAs, which release abnormal levels of Ca2+ signaling and overstimulation of glutamate. This overstimulation results in excitotoxicity, synaptic dysfunction, neuronal cell death, and a decline in cognitive functions. In clinical trials, several NMDA non-competitive antagonists have been tested, but most of them have failed due to low efficacy and the major concern of the side effects, which worsen the disease.[1,6] The first drugs licensed for symptomatic treatment of AD were ChE-Is, but currently, four drugs (donepezil, memantine, galantamine, rivastigmine) are approved. Table 4 (see Annex) outlines the mechanism of current AChE inhibitors and NMDA receptor antagonist drugs for Alzheimer's disease.[6]

Potential of CRISPR/Cas9 for treating AD:

CRISPR/CAS9 is a potential treatment that could be used for AD by gene-editing the mutations that cause it, potentially clearing patients from the disease. CRISPR/Cas9 utilizes two components for gene editing, one being the Cas9 enzyme and single-guide RNA (sgRNA). The sgRNA recognizes the desired DNA sequence to be modified, and the Cas9 protein acts as a pair of scissors, breaking the double strands of DNA. After this break, two pathways can be chosen to repair it, one resulting in the inactivation of the gene and the other in the replacement of a mutated sequence with the correct one. Homology-directed repair (HDR) utilizes a donor DNA template to replace a mutation with a correct sequence. As non-homologous end joining (NHEJ) leads to premature stop codons and DNA frameshifts, resulting in gene inactivation.3 As shown in Figure 3, CRISPR/Cas9 enables precise gene editing using guided RNA and Cas9 protein, and is delivered using viral or non-viral vectors.

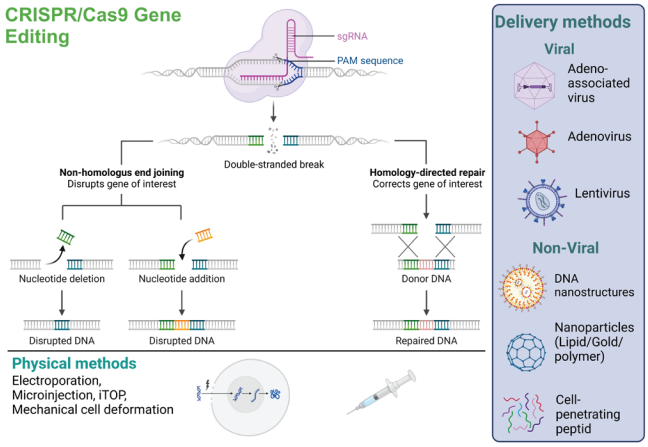


Figure 3: CRISPR/Cas9 gene editing tool. This figure illustrates the gene editing mechanism of CRISPR/Cas9 and the methods used for its delivery into cells, including viral and non-viral vectors. (Created by author).[10]

Treatment with CRISPR/Cas9 might not be as viable as other treatments because most AD cases are sporadic, and they involve unknown causes. Only less than 1% of familial cases of AD (FAD) account for genetic mutations and have negligible effects in sporadic AD (SAD). However, CRISPR/Cas9 might be considered as a therapeutic option to regulate beta-amyloid metabolism in FAD and SAD, reducing the overproduction of beta-amyloid, reducing the progression of the disease, or stopping it.[3] Personalized CRISPR has been used successfully to treat sickle cell disease (SCD) by gene editing hematopoietic stem cells.[11] Although possible, the major challenges for CRISPR/Cas9 in treating both AD and SCD are the low editing efficiency and the high off-target. However, if these challenges for CRISPR/Cas9 are resolved, the application of this tool could open new pathways for curing various diseases, such as AD and SCD.[12] CRISPR/Cas9 opens a pathway to reduce the progression or stop AD, but does not prevent it, by targeting the mutations that cause the accumulation of beta-amyloid or Tau protein. The modification of these genes comes with challenges that are discussed in this paper, as the methods of delivery of CRISPR/Cas9. To target the mutations that cause AD, a guide RNA that specifically binds to the region of DNA containing the mutation needs to be created. The codon where the mutation is can be corrected with HDR or deactivated the mutation with NHEJ.

Limitations of CRISPR/Cas9:

CRISPR/Cas9 lacks efficient and safe delivery methods to date. Currently, viral and non-viral methodologies are used for the delivery of CRISPR/Cas9. For viral methodologies, a concerning point is that they may incorporate mutations that have adverse effects on patients.

Viral methodologies:

Adeno-associated viruses (AAV) are frequently used vectors because of their high infectivity and low integration into the human genome, addressing the problem but lacking in efficiency due to the small amounts of genetic code that can be modified. Lentivirus is another type of virus that incorporates

DOI: 10.36838/v7i10.60

long DNA inserts, allowing bigger modifications but with lower brain dissemination efficiency. The negative factor for lentiviruses is the difficulty in the production of large quantities, and they cannot be incorporated into the human genome as easily as AAV, provoking immune reactions that are not desired.[3] Adenoviruses can't trespass the blood-brain barrier (BBB) by their size; on the other hand, non-viral methodologies are preferred to target the brain as they are smaller and easier to deliver.

Non-viral methodologies:

Non-viral methodologies are more promising to deliver CRISPR/Cas9 due to better cost-effectiveness and flexibility. They address the problem of mutations better because they are less immunogenic than viral vectors, which predispose to more immunogenic reactions. The formation of nanocomplexes by combining positively charged CRISPR/Cas9 peptides with negatively charged nucleic acid cargo doesn't present a huge challenge for the non-viral methodologies. The problem arises in the delivery methods of nanocomplexes in the brain, as they are unable to cross the blood-brain barrier (BBB) that protects the CNS from chemical substances in the blood and limits the chance of therapeutic drugs.[6] Another important factor is the reticuloendothelial system (RES), which actively removes nanocomplexes from the blood. Therefore, other delivery methods are used for nanocomplexes, such as intracerebroventricular and intrathecal injections. These direct infection methods require multiple injections for the proper distribution of nanocomplexes across the brain, which results in a complicated application.[3] Given current technology, it remains very difficult to utilize viral and non-viral methodologies to deliver CRISPR/Cas9 into the brain. There is a need to develop better technology that can facilitate the application of non-viral vectors, presenting a promising treatment for AD.

Conclusion

CRISPR/Cas9 is a novel tool that allows gene editing of specific gene sequences. It allows the elimination or replacement of unwanted mutations in genes that contribute to the development of diseases. CRISPR/Cas9 is present as a potential treatment for all diseases that have mutations in their genes involved such as FAD, a type of AD. AD lacks a cure, and the current treatments used for it only center on improving the cognition that is affected by AD. Independent of genetic factors in SAD and FAD, an altered amyloid beta metabolism is present, which leads to the accumulation of amyloid plaque in the brain that reduces synaptic function and causes the progressive deterioration of the brain. The causes of AD are multifactorial and mainly not related to genes. Mutations that provoke AD are only present in FAD, which represents 6% of all AD cases. CRISPR/Cas9 allows the modification of these mutations, presenting them as a potential treatment for 6% of AD cases, and establishing them as a new treatment that will clear the disease in all the patients who suffer from FAD. CRISPR/Cas9 could edit mutations that are linked with the development of FAD, such as mutations in the APP, PSEN-1, and PSEN-2 genes. CRISPR/Cas9 faces challenges such as efficient and safe delivery methods to edit mutations in humans. Viral and non-viral vectors are used to deliver CRISPR/Cas9 for potential therapy use in humans. Non-viral vectors are preferred because of their easier delivery methods and safer interactions with the receptor system, which do not include immunogenic reactions as viral vectors do. The main challenge of non-viral vectors is their small capacity to store genetic material and deliver it correctly, resulting in an inefficient therapy to cure the disease. For the future it is highly important to resolve the challenges that present for CRISPR/Cas9 when gene-editing for AD, the resolution of the challenges that limit this therapy will bring an effective treatment not only for AD but also for other diseases due that the challenges that present with CRISPR/Cas9 are not exclusive to AD, and present in general when gene-editing modification is applied for human therapies.

Acknowledgments

The author thanks Hamidreza Shaye, Lauren Tetz, and the IRIS program for their guidance in the research process.

References

1. Breijyeh, Z., & Karaman, R. (2020). Comprehensive Review on Alzheimer's Disease: Causes and Treatment. Molecules (Basel, Switzerland), 25(24), 5789.
2. Created in BioRender. Suárez, I. (2025) https://BioRender.com/t67p857
3. Bhardwaj, S., Kesari, K. K., Rachamalla, M., Mani, S., Ashraf, G. M., Jha, S. K., Kumar, P., Ambasta, R. K., Dureja, H., Devkota, H. P., Gupta, G., Chellappan, D. K., Singh, S. K., Dua, K., Ruokolainen, J., Kamal, M. A., Ojha, S., & Jha, N. K. (2022). CRISPR/Cas9 gene editing: New hope for Alzheimer's disease therapeutics. Journal of Advanced Research, 40, 207–221.
4. Marucci, G., Buccioni, M., Ben, D. D., Lambertucci, C., Volpini, R., & Amenta, F. (2021). Efficacy of acetylcholinesterase inhibitors in Alzheimer's disease. Neuropharmacology, 190, 108352.
5. Pandey, G., & Ramakrishnan, V. (2020). Invasive and non-invasive therapies for Alzheimer's disease and other amyloidosis. Biophysical Reviews, 12(5), 1175–1186.
6. 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.
7. Created in BioRender. Suárez, I. (2025) https://BioRender.com/n98p153
8. Pizzo, P., Basso, E., Filadi, R., Greotti, E., Leparulo, A., Pendin, D., Redolfi, N., Rossini, M., Vajente, N., Pozzan, T., & Fasolato, C. (2020). Presenilin-2 and Calcium Handling: Molecules, Organelles, Cells and Brain Networks. Cells, 9(10), 2166.
9. Lin, H., Li, G., Peng, X., Deng, A., Ye, L., Shi, L., Wang, T., & He, J. (2021). The Use of CRISPR/Cas9 as a Tool to Study Human Infectious Viruses. Frontiers in Cellular and Infection Microbiology, 11, 590989.
10. Created in BioRender. Suárez, I. (2025) https://BioRender.com/w86f686
11. Singh, A., Irfan, H., Fatima, E., Nazir, Z., Verma, A., & Akilimali, A. (2024). Revolutionary breakthrough: FDA approves CASGEVY, the first CRISPR/Cas9 gene therapy for sickle cell disease. Annals of medicine and surgery (2012), 86(8), 4555–4559.
12. Park, S. H., & Bao, G. (2021). CRISPR/Cas9 gene editing for curing sickle cell disease. Transfusion and Apheresis Science: Official

Journal of the World Apheresis Association: Official Journal of the European Society for Haemapheresis, 60(1), 103060.
13. Andrade-Guerrero, J., Santiago-Balmaseda, A., Jeronimo-Aguilar, P., Vargas-Rodríguez, I., Cadena-Suárez, A. R., Sánchez-Garibay, C., Pozo-Molina, G., Méndez-Catalá, C. F., Cardenas-Aguayo, M.-D.-C., Diaz-Cintra, S., Pacheco-Herrero, M., Luna-Muñoz, J., & Soto-Rojas, L. O. (2023). Alzheimer's Disease: An Updated Overview of Its Genetics. International Journal of Molecular Sciences, 24(4), 3754.
14. Arber, C., Lovejoy, C., Harris, L., Willumsen, N., Alatza, A., Casey, J. M., Lines, G., Kerins, C., Mueller, A. K., Zetterberg, H., Hardy, J., Ryan, N. S., Fox, N. C., Lashley, T., & Wray, S. (2021). Familial Alzheimer's Disease Mutations in PSEN1 Lead to Premature Human Stem Cell Neurogenesis. Cell Reports, 34(2), 108615.
15. Asanomi, Y., Kimura, T., Shimoda, N., Shigemizu, D., Niida, S., & Ozaki, K. (2024). CRISPR/Cas9-mediated knock-in cells of the late-onset Alzheimer's disease-risk variant, SHARPIN G186R, reveal reduced NF-κB pathway and accelerated Aβ secretion. Journal of Human Genetics, 69(5), 171–176.
16. Buccellato, F. R., D'Anca, M., Tartaglia, G. M., Del Fabbro, M., Scarpini, E., & Galimberti, D. (2023). Treatment of Alzheimer's Disease: Beyond Symptomatic Therapies. International Journal of Molecular Sciences, 24(18), 13900.
17. Devkota, S., Zhou, R., Nagarajan, V., Maesako, M., Do, H., Noorani, A., Overmeyer, C., Bhattarai, S., Douglas, J. T., Saraf, A., Miao, Y., Ackley, B. D., Shi, Y., & Wolfe, M. S. (2024). Familial Alzheimer mutations stabilize synaptotoxic γ-secretase-substrate complexes. Cell Reports, 43(2), 113761.
18. Han, J. P., Kim, M., Choi, B. S., Lee, J. H., Lee, G. S., Jeong, M., Lee, Y., Kim, E.-A., Oh, H.-K., Go, N., Lee, H., Lee, K. J., Kim, U. G., Lee, J. Y., Kim, S., Chang, J., Lee, H., Song, D. W., & Yeom, S. C. (2022). In vivo delivery of CRISPR-Cas9 using lipid nanoparticles enables antithrombin gene editing for sustainable hemophilia A and B therapy. Science Advances, 8(3), eabj6901.
19. Jucker, M., & Walker, L. C. (2023). Alzheimer's disease: From immunotherapy to immunoprevention. Cell, 186(20), 4260–4270.
20. Melancon, B. J., Tarr, J. C., Panarese, J. D., Wood, M. R., & Lindsley, C. W. (2013). Allosteric modulation of the M1 muscarinic acetylcholine receptor: Improving cognition and a potential treatment for schizophrenia and Alzheimer's disease. Drug Discovery Today, 18(23–24), 1185–1199.
21. Miller, M. B., Huang, A. Y., Kim, J., Zhou, Z., Kirkham, S. L., Maury, E. A., Ziegenfuss, J. S., Reed, H. C., Neil, J. E., Rento, L., Ryu, S. C., Ma, C. C., Luquette, L. J., Ames, H. M., Oakley, D. H., Frosch, M. P., Hyman, B. T., Lodato, M. A., Lee, E. A., & Walsh, C. A. (2022). Somatic genomic changes in single Alzheimer's disease neurons. Nature, 604(7907), 714–722.
22. Nithianandam, V., Bukhari, H., Leventhal, M. J., Battaglia, R. A., Dong, X., Fraenkel, E., & Feany, M. B. (2023). Integrative analysis reveals a conserved role for the amyloid precursor protein in proteostasis during aging. Nature Communications, 14(1), 7034.
23. Passeri, E., Elkhoury, K., Morsink, M., Broersen, K., Linder, M., Tamayol, A., Malaplate, C., Yen, F. T., & Arab-Tehrany, E. (2022). Alzheimer's Disease: Treatment Strategies and Their Limitations. International Journal of Molecular Sciences, 23(22), 13954.
24. Rossi S (2006) Australian medicines handbook. Australian Medicines Handbook, Adelaide, pp 2–3
25. Ruetz, T. J., Pogson, A. N., Kashiwagi, C. M., Gagnon, S. D., Morton, B., Sun, E. D., Na, J., Yeo, R. W., Leeman, D. S., Morgens, D. W., Tsui, C. K., Li, A., Bassik, M. C., & Brunet, A. (2024). CRISPR–Cas9 screens reveal regulators of ageing in neural stem cells. Nature, 634(8036), 1150–1159.
26. Shivnaraine, R. V., Kelly, B., Sankar, K. S., Redka, D. S., Han, Y. R., Huang, F., Elmslie, G., Pinto, D., Li, Y., Rocheleau, J. V., Gradinaru, C. C., Ellis, J., & Wells, J. W. (2016). Allosteric modulation in monomers and oligomers of a G protein-coupled receptor. eLife, 5, e11685.
27. Stefaniak, O., Dobrzyńska, M., Drzymała-Czyż, S., & Przysławski, J. (2022). Diet in the Prevention of Alzheimer's Disease: Current Knowledge and Future Research Requirements. Nutrients, 14(21), 4564.
28. Stefanoska, K., Gajwani, M., Tan, A. R. P., Ahel, H. I., Asih, P. R., Volkerling, A., Poljak, A., & Ittner, A. (n.d.). Alzheimer's disease: Ablating single master site abolishes tau hyperphosphorylation. Science Advances, 8(27), eabl8809.
29. Stephens, C. J., Lauron, E. J., Kashentseva, E., Lu, Z. H., Yokoyama, W. M., & Curiel, D. T. (2019). Long-term correction of hemophilia B using adenoviral delivery of CRISPR/Cas9. Journal of Controlled Release, 298, 128–141.
30. Su-Tobon, Q., Fan, J., Goldstein, M., Feeney, K., Ren, H., Autissier, P., Wang, P., Huang, Y., Mohanty, U., & Niu, J. (2025). CRISPR-Hybrid: A CRISPR-Mediated Intracellular Directed Evolution Platform for RNA Aptamers. Nature Communications, 16(1), 595.
31. Wu, W., Zhao, Y., Cheng, X., Xie, X., Zeng, Y., Tao, Q., Yang, Y., Xiao, C., Zhang, Z., Pang, J., Jin, J., He, H., Lin, Y., Li, B., Ma, J., Ye, X., & Lin, W.-J. (2025). Modulation of the glymphatic system by visual circuit activation alleviates memory impairment and apathy in a mouse model of Alzheimer's disease. Nature Communications, 16(1), 63. 32. Yang, Y., Bagyinszky, E., & An, S. S. A. (2023). Presenilin-1 (PSEN1) Mutations: Clinical Phenotypes beyond Alzheimer's Disease. International Journal of Molecular Sciences, 24(9), 8417

Authors

Iván, a sophomore at Tecnológico de Monterrey with an interest in biomedical engineering.

Annex

Table 1: Variants in APP associated with familial Alzheimer's disease.
This table lists key mutations in the APP gene associated with familial Alzheimer's disease. Several variants, such as the Swedish and Flemish mutations, increase amyloid-beta production or aggregation, contributing to early disease onset.

Variant ID	Position	Amino Acid Change	Clinical Significance	dbSNP	Citations (Substitute SNP in link with dbSNP) (https://www.ncbi.nlm.nih.gov/snp/)
VAR_000015	670-671	KM>NL	In AD1; Swedish mutation; highly increases hydrolysis by BACE1 and amyloid-beta protein production	rs28186516	Yes
VAR_044424	678	D>N	In AD1	rs63750064	Yes
VAR_000016	692	A>G	In AD1; Flemish mutation; increases the solubility of processed amyloid-beta peptides and the stability of peptide oligomers	rs63750671	Yes

DOI: 10.36838/v7i10.60

VAR_014215	693	E>G	In AD1	rs63751039	Yes
VAR_000019	713	A>T	In AD1	rs63750066	Yes
VAR_032277	714	T>A	In AD1	rs63750643	Yes
VAR_014218	714	T>I	In AD1; increased amyloid-beta protein 42/40 ratio	rs63750973	Yes
VAR_010108	715	V>M	In AD1; decreased amyloid-beta protein 40/total amyloid-beta	rs63750734	Yes
VAR_000020	716	I>V	In AD1	rs63750399	Yes
VAR_000023	717	V>F	In AD1; increased amyloid-beta protein 42/40 ratio	rs63750264	Yes
VAR_000022	717	V>G	In AD1; increased amyloid-beta protein 42/40 ratio	rs63749964	Yes
VAR_000021	717	V>I	In AD1; increased amyloid-beta protein 42/40 ratio	rs63750264	Yes
VAR_014219	717	V>L	In AD1	rs63750264	Yes
VAR_010109	723	L>P	In AD1	rs63751122	Yes

Table 2: Variants in PSEN 1 are associated with familial Alzheimer's disease.
This table summarizes mutations in the PSEN1 gene linked to familial Alzheimer's disease. Many of these variants increase the Aβ42/Aβ40 ratio, disrupt protease activity, and are highly pathogenic.

Variant ID	Position	Amino Acid Change	Clinical Significance	dbSNP	Citations (Substitute SNP in link with dbSNP) (https://www.ncbi.nlm.nih.gov/snp/)
VAR_075260	35	R>Q	In AD3; uncertain significance; decreased protease activity with APP	rs63750592	Publications
VAR_006413	79	A>V	In AD3; also found in late-onset Alzheimer disease; impaired protease activity with APP; altered amyloid-beta production and increased amyloid-beta 42/amyloid-beta 40 ratio; no effect on GFAP	rs63749824	Publications
VAR_006414	82	V>L	In AD3; decreased protease activity with APP; no effect on GFAP	rs63749967	Publications
VAR_075261	83	I>T	In AD3	-	1 Publication
VAR_081228	85	L>P	In AD3; spastic paraparesis and apraxia; loss of protease activity with APP; altered amyloid-beta production and increased amyloid-beta 42/amyloid-beta 40 ratio	rs63750599	Publications
VAR_081229	89	V>L	In AD3; decreased protease activity with APP; increased amyloid-beta 42/amyloid-beta 40 ratio	rs63750815	Publications
VAR_016214	92	C>S	In AD3; loss of protease activity with APP	rs63751141	Publications
VAR_081230	94	V>M	In AD3; uncertain significance; reduced protease activity with APP; no change in amyloid-beta 42/amyloid-beta 40 ratio	rs63750831	Publications

Table 3: Variants in PSEN 2 associated with familial Alzheimer's disease.
This table outlines rare PSEN 2 mutations associated with familial Alzheimer's disease. Some of these variants affect calcium regulation and amyloid-beta production.

Variant ID	Position	Amino Acid Change	Clinical Significance	dbSNP	Citations (Substitute SNP in link with dbSNP) (https://www.ncbi.nlm.nih.gov/snp/)
VAR_006461	62	R>H	In AD4; likely benign	rs58973334	Publications
VAR_070027	71	R>W	In AD4; uncertain significance	rs140501902	Publications
VAR_009214	122	T>P	In AD4	rs63749851	Publication
VAR_081261	122	T>R	In AD4; increased mitochondrion-endoplasmic reticulum membrane tethering resulting in increased calcium transfer to mitochondria	rs28936380	Publications
VAR_081262	126	E>K	In AD4; uncertain significance	-	1 Publication
VAR_064903	130	S>L	In CMD1V and AD4; uncertain significance	rs63750197	Publications
VAR_006462	141	N>I	In AD4; results in altered amyloid-beta production and increased amyloid-beta 42/amyloid-beta 40 ratio; loss of function as calcium-leak channel;	rs63750215	Publications

			calcium overload in the ER		
VAR_081263	141	N>Y	In AD4	rs61761208	Publication
VAR_007958	148	V>I	In AD4; late-onset Alzheimer disease	rs63750812	Publication
VAR_009215	239	M>I	In AD4	rs63749884	Publication
VAR_006463	239	M>V	In AD4	rs28936379	Publication

Table 4: Drugs used for Alzheimer's disease.
This table summarizes the mechanisms of current treatments that focus on symptom management rather than disease modification.

Drug Name	Drug Type	Target	Mechanism/Effect
Donepezil	Acetylcholinesterase Inhibitor	Acetylcholinesterase (enzyme)	Inhibits acetylcholinesterase, increasing acetylcholine levels in the synaptic cleft; enhances cholinergic function.
Memantine	NMDA Receptor Antagonist	NMDA Receptor (N-methyl-D-aspartate receptor)	Blocks excessive glutamate activity at NMDA receptors, preventing excitotoxicity associated with neurodegeneration.
Galantamine	Acetylcholinesterase Inhibitor	Acetylcholinesterase; Nicotinic receptors (allosterically)	Inhibits acetylcholinesterase and modulates nicotinic receptors, increasing acetylcholine and enhancing neurotransmission.
Rivastigmine	Acetylcholinesterase and Butyrylcholinesterase Inhibitor	Acetylcholinesterase and Butyrylcholinesterase (enzymes)	Inhibits both acetylcholinesterase and butyrylcholinesterase, increasing acetylcholine levels to improve cognition and behavior.

DOI: 10.36838/v7i10.60

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

The Effects of Regional Environments on Entrepreneurial Intentions: Exploring Global Entrepreneurship

Gaod Dagher

The International School of Choueifat- Amman,11810 Airport Rd., Amman, Jordan judedagher@gmail.com

ABSTRACT: Entrepreneurship is a skill that has been around for centuries and has been practiced in various ways throughout the years. After the COVID-19 pandemic, more people have become interested in becoming entrepreneurs. As defined by Bird (1988), entrepreneurial intention is "an entrepreneur's state of mind that directs attention, experience, and action towards a business concept." This paper aims to explore the various environmental factors that might affect a person's entrepreneurial intent. By examining people with entrepreneurial intentions in different countries in the Middle East*, patterns emerge regarding economic stability and development in various countries in the Region. Considering the fact that traditional entrepreneurship education does not usually expose students to the entrepreneurial process through problem-solving, entrepreneurial education has not had major success in helping students start business ventures. Looking at entrepreneurs in the Middle East, this paper will attempt to explore how an individual's environment shapes the entrepreneurial mindset. This will help several institutions and scholars determine if further efforts should be implemented to develop future entrepreneurs.

KEYWORDS: Behavioral and Social Sciences, Sociology, Entrepreneurial Intentions, Regional Environments, Entrepreneurship Education, Middle East.

Introduction

Entrepreneurship is a driving force to economic growth and success, and while entrepreneurship has been around for centuries, the term 'entrepreneur' is defined differently by many scholars. For the sake of this research, an entrepreneur is any individual with entrepreneurial intent or the intention of starting a business.[1] Is there a certain psychological aspect that drives a person to become an entrepreneur?[2] Research suggests there is a direct link between personality traits and entrepreneurial efficacy.[3] Other factors include an entrepreneur's environment; living in economically prosperous countries with financial stability is substantially different than living in a developing country riddled with chaos. The need to rise above poverty must also be considered; when faced with impoverishment, a person might turn to take their fate into their own hands and aim to start their business solely to survive.[3] Of course, personality traits are largely influenced by the cultural beliefs a person grows up around, which transitively means that personality traits are affected by regional factors. After establishing this point, it becomes clear that a person's psyche is directly linked to their environment, which both affect entrepreneurial intent. Entrepreneurship is associated with the ability to take risks and inventiveness, which suggests that entrepreneurship is not a default skill in the human mind, but a specialty that some possess, which can be nurtured and grown. Another question that raises itself is 'can entrepreneurial education enhance entrepreneurial intention in persons with the entrepreneurial seed planted within them because of their upbringing?'

Taking an entrepreneurship course is undoubtedly a better option than taking a leap of faith and becoming an entrepreneur without any guidance, but one cannot be sure of just how effective it is in comparison to regional conditions. Entrepreneurial learning relies heavily on a linear approach of problem-solving, which lacks the suitable tools to solidify a student's contextual understanding of the entrepreneurial process.[4] The Middle East is a region with many developing countries and unstable economies, so understanding the effectiveness of entrepreneurial programs along with the core drivers for entrepreneurial intent would help them encourage entrepreneurship more effectively, thus kickstarting positive changes to the economy. Several countries in the Middle East, specifically the Gulf region, have launched entrepreneurial education systems with the motive of gaining a competitive global advantage. The UAE, in particular, has launched a program to utilize undergraduates' ideas into businesses, providing a community for students [3] to have access to key stakeholders and businessmen, building a collaborative environment that helps college students identify key phases of the entrepreneurial process including 'ideation', 'incubation', and 'acceleration'.[4] The Gulf region has always had a strong stance in wanting to develop future entrepreneurs, [1] and often puts resources into the matter. Establishing entrepreneurship programs is essential to the prosperity of the Middle East, as entrepreneurship often spurs innovation and economic growth.

While some scholars think that reshaping the entrepreneurial education system will result in fostering a more creative process, with more innovation and research skills, that does not necessarily mean that more people will develop entrepreneurial intent. Training students to work more collaboratively, giving them tools to create new ideas, and teaching them to analyze information to make quick decisions has an overflow of benefits to employability and economic prosperity, yet factors such as the need for achievement (nAch) and locus of control (LoC) rely solely on a person's psychological traits and cannot be changed by learning mechanisms.[3] In a region as variable

DOI: 10.36838/v7i10.67

as the Middle East, one can determine the effect of political backgrounds and economies on entrepreneurial intent. The Middle East ranges from some of the richest to some of the poorest countries in the world, which provides the opportunity to analyze how the risk of poverty affects an entrepreneur's actions. This study aims to show the differences a person's environment makes on their entrepreneurial intent.

Literature Review

Entrepreneurial programs often have a positive impact on thinking, communication, and time-management skills.[2] Entrepreneurship education provides a space for applying divergent creativity thinking tools that enable students to acquire competencies to secure employment opportunities and have life-long skills of value; having entrepreneurial education based on utilizing divergent creativity processes will undoubtedly emerge with confidence in creativity skills that allow students to generate advantageous ideas, maximize idea breadths, originality and the ability to identify gaps, complexity of schemes, and curiosity about different ventures, among other skills.[4–6] While all these advantages may seem appealing to anybody, these programs will not affect a person with no underlying hunger for entrepreneurship. No learning program can magically grow the nAch and LOC in a person, no matter how effective it is.[3]

The Middle East, particularly, is a useful region in identifying the effect of background and upbringing of a person in making them an entrepreneur. In the Middle East, countries like Yemen and Syria are still in an economic crisis and have governmental chaos. On the other hand, countries of the Gulf are considered relatively safe and stable. So that raises the question: how does that affect a person's entrepreneurial mindset? Is desperation a motivator for entrepreneurial intent, or is someone living a safe and stable life more likely to start a business? Lately, entrepreneurship has stopped becoming singular to one person, with the focus shifting to entrepreneurial ecosystems, and not only a person's individual characteristics and personality.[7] The concept of regional social and economic environments affecting entrepreneurial growth and intention is not new by any means, with the work of Spigel, Malecki, and Ritsilä highlighting the geographical perspective.[8–10]

To understand how regional differences impact entrepreneurship, one must analyze the economic standards in different countries in the Middle East:

Changes in household income throughout different countries:

As shown in Figure 1, this figure demonstrates that, compared to Level B or Level A economies (relatively prosperous economies), low-income economies were significantly more likely to have seen a decline in household income in 2023. For instance, compared to only four out of seventeen Level B economies and none of the Level A economies, adults in eleven out of fourteen Level C economies reported that their household income had decreased in the previous year by at least 40%.11: When a person witnesses significant decreases in their household income from traditional wage-earning methods, they might think of turning to starting businesses to improve their living standards.

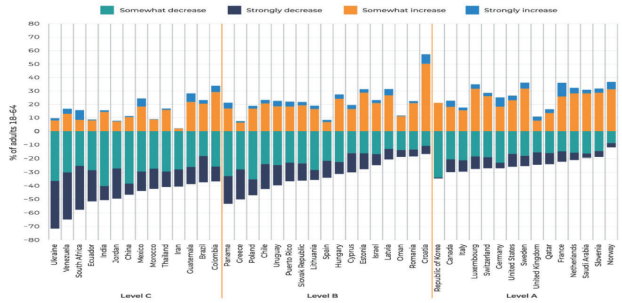


Figure 1: Graphical representation of differences in household incomes throughout different economies globally, with many different classes of Middle Eastern economies. Low-income economies (Level C) experienced significant declines in household income, while Levels A and B showed relative stability or growth. Ukraine recorded the highest percentage of households reporting income losses, whereas Croatia and Norway had the highest levels of income growth.[12]

As this graph suggests, Figure 1, that Level C economies (developing countries) may have more incentives for business startups, as conventional labor has caused significant decreases in their earnings. Starting a business would start to look very appealing to anybody looking to increase their income.

Entrepreneurial activity in different countries:

Figure 2 displays the TEA and EBO levels for each of the 46 economies. In four Latin American and Caribbean economies—Guatemala, Ecuador, Chile, and Panama—more than three out of ten adults were launching or operating a new company. With a few exceptions in Latin America, levels of early-stage entrepreneurial activity generally declined as income increased. There were five Level C economies, with one in five or more adults starting or running a new business, compared to four in Level B and just one in Level A. In contrast to Levels B and A combined, which had 13 economies out of 32, Level C had only three economies out of 14, with less than 10% of adults launching or operating a new business.[11] This graph demonstrates how popular business startups are in Level C economies, cementing the point suggested by Figure 1.

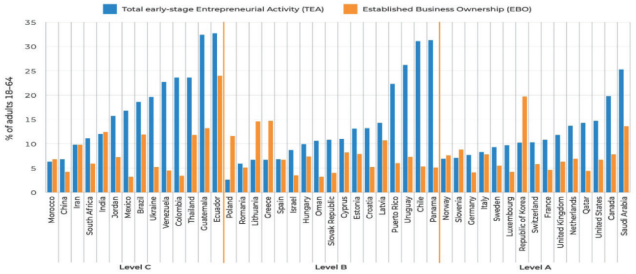


Figure 2: Total early-stage entrepreneurial activity (TEA) and established business ownership (EBO) (both % adults). Levels of TEA and EBO across 46 economies, early stage entrepreneurial activity (TEA) was highest in Latin America & Carribean, with general decline as income increased, while established business ownership (EBO) showed less variation and, in some Level C economies, matched or exceeded TEA levels.[12]

In these graphs (Figures 1-2), countries are separated into levels A, B, and C which separates economies into three lev-

els: level C refers to a level of GDP per capita of less than $25,000, level B has a GDP per capita of $25,000-$50,000, level A has a GDP per capita of more than $50,000. Middle Eastern countries are included in all three categories, showing that early-stage entrepreneurial activity is more popular than established business ownership.

Statistics suggest that richer economies in the Middle East have a higher business start-up rate, which is thanks to programs established to create a safe environment for entrepreneurs to prosper. By using results from a study by Al Qadasi *et al.* (2023), which found that the need for achievement and entrepreneurial self-efficacy correlate with political environments, one can infer that unstable environments might create a need for business start-ups, but they still need support from experienced individuals.[3]

One very notable thing is that while most Middle Eastern countries dropped in entrepreneurial activity after 2019, Saudi Arabia had a significant increase in 2023, which might be because of the Monsha'at program for entrepreneurship, which partnered with Silicon Valley and MIT in 2023, clearly having an impact on entrepreneurial activity. Qatar also had an increase in entrepreneurial activity in 2021, then dropped back in 2023. The two graphs above show the economic differences in different countries in the Middle East, which gives insight into whether entrepreneurship programs in said countries have helped boost their economic growth and entrepreneurial activity.

As Al Qadasi *et al.* have previously proven in Ye-men, socio-political situations can heavily impact a person's entrepreneurial intent; in an unstable environment with high unemployment rates, a business start-up is viewed as the best source of income.[3] If this hypothesis can be proven in Yemen, it can be proven in many other Middle Eastern countries with similar circumstances, such as Jordan, with a staggering 40.80% rate of unemployment among youth in 2023.[13] Al Qadasi et al. have shown that attitudes, personality traits, and contextual factors influence an individual's entrepreneurial intent.[3]

Environmental variables interact with a person's personality traits to affect entrepreneurial behavior. Entrepreneurial ecosystems are crucial for establishing entrepreneurial social networks, which could help motivate and fund an entrepreneur's business start-up. Entrepreneurial intent in such countries is "necessity-based" and shows correlations between personality traits and situational factors that affect entrepreneurial intent.

Methodology

This paper combines literature about how environmental factors combine with personality traits that might affect entrepreneurial intent. By analyzing research conducted in the Middle East. This study provides a detailed analysis of the different economic situations in the Middle East to help the reader understand what affects entrepreneurial intent. For developing economies in the Middle East, entrepreneurship is a much-needed boost for unemployment rates, and the only way to nurture entrepreneurship is to identify the factors that push entrepreneurial intent.

Most literature has been derived from Google Scholar, between the time period of 2018 to 2024, only including research from before that period if it's necessary to build a foundation. After analyzing data on the effects of the environment on entrepreneurial intent, there was also a connection between entrepreneurship education, which means that the effectiveness of entrepreneurial education is vital to inspiring an entrepreneur to start a business. After evaluating the correlation between socio-political environments, unemployment, and entrepreneurship education, it was found that this paper needed to include all three factors in order to have a clear study.

This review puts together multiple different perspectives on the topic, in order to give the reader an easy introduction to the world of entrepreneurial intentions. The research provides a brief introduction to the effect of environments on entrepreneurial intent, proving the significant effect a person's environment has on psychology, which in turn affects entrepreneurial intentions. It is recommended to read the cited sources as well as other literature for a more in-depth understanding of all aspects that affect entrepreneurial intent. The limitations of this study include the need for a more comprehensive inquiry around testing the effectiveness of entrepreneurship education with persons of entrepreneurial intent in more chaotic economies, and whether entrepreneurship education, as it stands today, is effective in building entrepreneurial intent. Further research needs to be conducted on necessity and opportunity-based entrepreneurial intent in different environments. Suggestions for future research also include testing the effectiveness of entrepreneurial ecosystems in developing Middle Eastern countries. There is a need for more detailed research on the effect of a person's family history and to expand on looking into the psychological aspects affecting entrepreneurial intent regionally.

Results

Psychology plays a huge part in entrepreneurial success, with results correlating positively between locus of control and entrepreneurial ability. This psychology is largely affected by upbringing and cultural standards. Several environmental factors affect the extent of entrepreneurial intent a person has, including living conditions. This shows that entrepreneurship cannot be taught, but can be nurtured instead. Current courses do not focus on the psychology behind entrepreneurial intent but on economic standards regarding entrepreneurship.

A person with no previous interest in entrepreneurship will not benefit from said courses. It is still unclear if entrepreneurial intent can be taught, as several scholars disagree. Entrepreneurship greatly impacts economic growth and development, specifically in higher-income countries.

The "entrepreneurial economy" is a popular method that economies use to improve unemployment problems and enrich their economy.[14] Highly developed economies tend to tap into fostering entrepreneurship more than underdeveloped countries, especially in the Middle East, which helps them get higher entrepreneurship rates, although based on previous studies, countries with economic voids have great potential to stabilize their economic status by fostering entrepreneur-

DOI: 10.36838/v7i10.67

ial ecosystems. Individuals with entrepreneurial intent tend to have higher self-efficacy and locus of control rate than others.[15] Entrepreneurial education can include personality traits and personalized courses based on environmental factors to encourage entrepreneurship in their country.

Entrepreneurial education may become more effective in encouraging business start-ups if the programs focus on creativity and situational factors. Entrepreneurial intent in developing countries in the Middle East is driven by necessity, while it is driven by opportunity in richer Middle Eastern countries.[3] Although countries like the UAE have launched several programs to encourage entrepreneurial activity and business ventures, only 2% of citizens have the desire to start a business, which is caused by the cost of failure, which can be monetary, social, or motivational. Individuals are afraid to seize the opportunities they have because they fear bringing shame and embarrassment to themselves, losing money, and losing their motivation after failing.[15]

Citizens of rich Middle Eastern countries recognize the opportunities they have, but are too afraid of taking the risk and failing. On the other hand, citizens of developing Middle Eastern countries have higher entrepreneurial intent but lack the opportunities and support required to start a business. Entrepreneurs who are business owners have a higher creativity rate than employees.[15]

All in all, all these observations reach one clear point: environments where low income is available provide a perfect breeding ground for entrepreneurial intent, as individuals have a larger need for achievement, but the resources provided by richer economies often give business start-ups a larger chance at succeeding.

Discussion

It is very important to note that this paper is merely an introduction to a topic that goes far deeper than the scope of this review. The author hopes that further research is conducted I this field with deeper explorations of cultural norms, internal environments, and individual beliefs. That being said, this paper has given rise to a threshold of research possibilities that explore this topic justly, and opens up opportunities for people to understand the extent of just how impactful their environment is to their lives.

Increasing attention has been given to entrepreneurs in developing countries, as it has been found that entrepreneurial intentions are a "regional dimension." [14] This is only the first step into something major, and we are severely lacking in research when it comes to entrepreneurial intentions regionally.

Throughout this study, the term entrepreneurial intention has been used in the context of an aware person having a conscious intention to develop a business start-up. Regional factors are not only a person's living standards, but also whether or not they have a family business, or live in a supportive household among other factors that can affect an individual's entrepreneurial intent. It has been found that environmental factors have an impact on a person's need for achievement, which in turn leads to entrepreneurial intent. With that knowledge, policymakers may invest in creating entrepreneurial ecosystems for university students to foster their entrepreneurial intent into a business start-up, which has been known to boost an economy's financial status.

Middle Eastern countries, as countries that have either developing economies or recently developed ones, served as the perfect ground to explore such a topic. Countries like Saudi Arabia, with high funding for entrepreneurial programs, have experienced high rates of business startups. On the other hand, countries like Libya, with not much focus on entrepreneurship, have an increasing rate of entrepreneurial intent despite not having the proper tools.

Having underdeveloped governments invest in entrepreneurial programs may also result in negative consequences if the educational programs are not personalized to the country's political and economic environment or individuals' personality traits. Financing new business start-ups is also essential to encourage new entrepreneurs to start businesses. Entrepreneurship has been proven to boost a country's economy, and in cases like Yemen and Libya, entrepreneurship would be a long-term solution to their current economic crisis.

The Gulf countries, particularly Saudi Arabia and the UAE, have heavily invested in building entrepreneurial environments that foster entrepreneurial intention into bigger opportunities and business ventures, which has allowed citizens to start their own businesses. Confidence-building exercises can also be implemented into entrepreneurial education courses to eliminate the fear of failure that accompanies starting a new business. After looking at entrepreneurship regionally, it is clear that environmental aspects affect entrepreneurial intent and the desire to start business ventures. Overall, entrepreneurial intent is affected by situational factors as well as psychological aspects.

This review would help policymakers and investors understand how the entrepreneurial mindset is built, in order to put forth programs that correctly nurture entrepreneurial environments. Individuals would also benefit from this paper by having a clear layout of how entrepreneurial intention is affected by different factors, and whether the current environment they're in positively or negatively impacts their entrepreneurial intent.

Conclusion

Further research may include looking at a person's family history regarding entrepreneurship, analyzing how the fear of failure hinders the desire to start a business, and looking at how entrepreneurship education can be improved.

This study compared how the environmental factors of different countries in the Middle East affect a person's need for achievement and entrepreneurial intent. It was found that a person in a developing country with challenging living circumstances is more likely to start a business than one in a comfortable living environment with no real need to improve their conditions. It was also found that developing countries can stabilize their economies by focusing on developing entrepreneurial intent into a long-term business start-up. Entrepreneurial ecosystems need to support an entrepreneur and guide them in making the right decisions. Without the right funding and guidance, governments will not be able to take advantage of entrepreneurial intent among youth. For a lasting

impact on entrepreneurial environments, entrepreneurs need to be allowed to thrive, especially in unstable environments.

*__Middle East__, the lands around the southern and eastern shores of the Mediterranean Sea, encompassing at least the Arabian Peninsula and, by some definitions, Iran, North Africa (Britannica Encyclopedia)

Acknowledgments

I attest that the ideas, graphics, and writing in this paper are entirely my own. I would like to thank Dr. Michellana Jester and Mr. Neil Bui for their constant guidance and helpfulness. I would never have gotten the confidence to write this paper without their support and supervision. A huge thank you to my mom and dad, who always strive to give me the opportunities to achieve my dreams. I would be nothing without both of you.

References

1. Saleh and Salhieh - 2014 - An Investigation of Entrepreneurial Intention amongst Arab University Students.Pdf. https://ccsenet.org/journal/index.php/ijbm/article/download/38244/23233 (accessed 2025-03-13).
2. Ismail, A.; Adnan, W. N.; Masek, A.; Hassan, R.; Hashim, S.; Ismail, & M. E. Effectiveness of Entrepreneurship Programmes in Developing Entrepreneurship Skills Towards Quality TVET Graduates. *jtet* **2019**, *11* (1).
3. Al-Qadasi, N.; Zhang, G.; Al-Awlaqi, M. A.; Alshebami, A. S.; Aamer, A. Factors Influencing Entrepreneurial Intention of University Students in Yemen: The Mediating Role of Entrepreneurial Self-Efficacy. *Front. Psychol.* **2023**, *14*, 1111934. https://doi.org/10.3389/fpsyg.2023.1111934.
4. Ghafar, A. Convergence between 21st Century Skills and Entrepreneurship Education in Higher Education Institutes. *IJHE* **2020**, *9* (1), 218. https://doi.org/10.5430/ijhe.v9n1p218.
5. Schmidt, J. J.; Soper, J. C.; Facca, T. M. Creativity in the Entrepreneurship Classroom; Jordan Whitney Enterprises, Inc, 2011; Vol. 17, p 9.
6. Yar Hamidi, D.; Wennberg, K.; Berglund, H. Creativity in Entrepreneurship Education. *Journal of Small Business and Enterprise Development* **2008**, *15* (2), 304–320. https://doi.org/10.1108/14626000810871691.
7. Coduras, A.; De La Vega, I.; Salman, M. B. Entrepreneurial Ecosystems in Arabic Countries: A Recent Overview. In *Entrepreneurship Ecosystem in the Middle East and North Africa (MENA)*; Faghih, N., Zali, M. R., Eds.; Contributions to Management Science; Springer International Publishing: Cham, 2018; pp 283–302. https://doi.org/10.1007/978-3- 319-75913-5_9.
8. Stam, E.; Spigel, B. *Entrepreneurial Ecosystems*; USE Discussion paper series, 2016.
9. Malecki, E. J. Technology and Economic Development: The Dynamics of Local, Regional, and National Change. Social Science Research Network: Rochester, NY 1997. https://papers.ssrn.com/abstract=1496226 (accessed 2025-03-13).
10. Ritsila, J. J. Regional Differences in Environments for Enterprises. *Entrepreneurship & Regional Development* **1999**, *11* (3), 187–202. https://doi.org/10.1080/089856299283164.
11. GEM 2023/2024 Global Report: 25 Years and Growing.
12. *GEM Global Entrepreneurship Monitor*. GEM Global Entrepreneurship Monitor. https://www.gemconsortium.org/report/global-entrepreneurship-monitor-gem-20232024-global-report-25-years-and-growing (accessed 2025-03-13).
13. *Overview*. World Bank. https://www.worldbank.org/en/country/jordan/overview (accessed 2025-03-13).
14. Saleh, W.; Salhieh, L. An Investigation of Entrepreneurial Intention amongst Arab University Students. *IJBM* **2014**, *9* (12), p197. https://doi.org/10.5539/ijbm.v9n12p197.
15. Eid, R.; Badewi, A.; Selim, H.; El-Gohary, H. Integrating and Extending Competing Intention Models to Understand the Entrepreneurial Intention of Senior University Students. *ET* **2019**, *61* (2), 234–254. https://doi.org/10.1108/ET-02-2018-0030.

Author

Gaod (Jude) Dagher is an Iraqi born in Jordan in 2010. From childhood, Jude has had a curious mind with a need to answer all questions that come to mind, which has led her to conduct research and write about it. Jude hopes to graduate with honors and lead a long and fulfilling career.

DOI: 10.36838/v7i10.67

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Overview of Rocket Engine Cycles and an Analysis of Parameters via Proptools Simulations

Katherine Wang

The Harker School, 500 Saratoga Ave, San Jose, CA, 95129, USA; katherinewang793@gmail.com

ABSTRACT: Rocket propulsion systems commonly utilize bi-propellant combustion cycles, including open-cycle, closed-cycle oxygen-rich, closed-cycle fuel-rich, and full-flow staged combustion, each tailored to specific performance and design requirements. This paper provides an in-depth review of each combustion cycle, outlining each system's performance features and applications. Additionally, the paper discusses potential future designs, such as nuclear propulsion and tri-propellent liquid propulsion. Rocket propulsion systems are developed to maximize the efficiency of the system, ensuring that the rocket can produce the most thrust for a certain amount of fuel. The primary goal of this study is to understand which relationships between design parameters most significantly contribute to the optimal efficiency of a propulsion system. Simulation analyses conducted in Python using the proptools package demonstrate the effects of parameters on specific impulse, mass flow rate, and thrust to optimize system performance. The efficiency (i.e., specific impulse) is most influenced by changes in chamber temperature and exit pressure. For example, increasing exit pressure from 100 kPa to 500 kPa can reduce specific impulse by over 10%. Additionally, changes in throat area and chamber temperature impact mass flow rate, while exit pressure and throat area influence rocket thrust.

KEYWORDS: Engineering Mechanics, Aerospace and Aeronautical Engineering, Rocket, Rocket Engine Cycles.

Introduction

In the 20th-21st century, rocket propulsion is a critical technology that enables space exploration and advances humanity's understanding of the universe. This research explores the principles and performance of bi-propellant combustion cycles, including open-cycle, closed-cycle (oxygen-rich and fuel-rich), and full-flow staged combustion. By using simulations with the proptools library,[1] we analyze the impact of key parameters such as chamber temperature, throat area, and exit pressure on specific impulse, thrust, and mass flow rate. These insights bridge theoretical understanding and practical applications, offering a foundation for developing more efficient and reliable propulsion systems to meet the evolving demands of space missions.

I: Fuel:

Most propulsion systems, specifically chemical propulsion systems, rely on the energy released from propellants.[2] This energy is converted into thrust, governed by the principles of fluid dynamics and thermodynamics.

To maximize the efficiency of propulsion systems, several key factors must be considered. High combustion chamber temperatures are crucial, as they directly correspond to the energy released during combustion. This is achieved by using propellants with high energy content, where molecules with lower bond energy disintegrate more easily, releasing energy more effectively.[2]

Another critical factor is the isentropic coefficient,[3] which measures the relationship between changes in pressure and volume during gas expansion. When there is a small isentropic coefficient, it leads to more efficient energy transfer within the exhaust system. Essentially, this means that the system is better at managing heat, allowing it to transfer energy more effectively and improve overall performance.[4] A smaller molecular weight in the exhaust gas mixture is also preferred because lighter molecules require less energy to accelerate, leading to higher nozzle speeds and more efficient propulsion.[5]

The propellant's density and temperature are also critical factors in determining engine performance. High-density propellants enable greater mass to be burned per unit volume,[6] increasing the energy output for a given engine size. The temperature of the propellant is significant as well, as low-temperature liquids (such as cryogenic fuels) require substantial insulation to avoid heat regression. If the propellant has a low condensation point, there would be a chance for the liquid fuel to turn into a gas, which would negatively impact the fuel's performance. Next, we discuss the 3 most common types of fuel in a bi-propellant rocket engine.

Bi-propellent engines use liquid oxygen (LOX) and a fuel for combustion.[7] One of the most commonly used fuels in such systems is RP-1, which stands for Rocket Propellant-1, or refined petroleum-1.[7] RP-1 is a type of refined kerosene, a hydrocarbon fuel that is chosen primarily for its higher density, lower specific impulse compared to the more widely used liquid hydrogen (LH2), and more stable storage characteristics at room temperature.[7] RP-1 presents less of an explosion hazard than LH2, making it easier to handle and store. The development of RP-1 traces back to early attempts with alcohols, but scientists found hydrocarbons like kerosene to be more efficient for rocket engines due to their lack of an oxygen atom and negligible water content.

However, early attempts to burn kerosene in rocket engines led to unmanageable engine temperatures of ~2000C°.[8] Raw kerosene polymerizes at high temperatures, causing break-

DOI: 10.36838/v7i10.72

downs in engine components, overheating of chamber walls, burn-through, and rapid failure. To address this issue, scientists developed RP-1, a heat-resistant hydrocarbon fuel, which helped prevent these issues, making it a viable option for use in spaceflight.

Recall that LH2 is commonly used for bi-propellant engines as well, providing the highest specific impulse of any rocket fuel; however, liquid hydrogen comes with significant challenges. For example, LH2 must be stored at extremely low temperatures, around -423˚F (20K),[9] due to its low boiling point. Its low density means that large quantities of liquid hydrogen must be used for efficient rocket propulsion, which necessitates larger storage tanks in rockets. Hydrogen storage also poses a problem in that it can cause hydrogen embrittlement, a phenomenon where metals become brittle and lose ductility when exposed to hydrogen, potentially weakening the rocket's structural integrity. LH2's clean combustion process, however, is a significant advantage, as it reacts with oxygen to produce only water vapor, making it environmentally friendly.[9]

Liquid methane (LCH4), a newer option for rocket fuel, is used in several modern propulsion systems, such as SpaceX's Raptor and Blue Origin's BE-4 engines.[10] LCH4 has a lower specific impulse than liquid hydrogen but is easier to store and handle because it has a higher boiling point and density,[11] providing an easier avenue for efficient storage. Additionally, LCH4 is a promising option for *in situ* resource utilization (ISRU) on Mars, where it can potentially be produced from local resources on the Martian surface.[12] Compared to RP-1, liquid methane also burns cleaner, leaving less residue, making it a more environmentally friendly option while still offering sufficient performance for many rocket stages.

II: Rocket cycles:

The basics of a rocket engine essentially follow Newton's third law: For every action, there is an equal and opposite reaction. As seen in Figure 1, these rocket engines work by shooting material from the back of the rocket, whether that be through a pressurized vessel or a combustion between an oxidizer and fuel. [13,14]

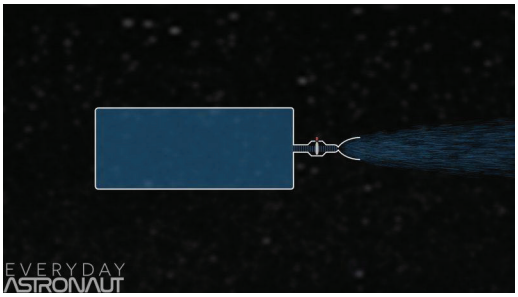


Figure 1: A simple demonstration of a rocket engine with a tank, valve, and nozzle. [10]

The simplest setup is a single-propellant engine that is just pressurized gas. When the valve of the tank is open, the gas flows out of the tank and into space, propelling the rocket forward. The more complex situation is running an engine with combustion, i.e., a bi-propellant engine, as mentioned in Section 1. This typically involves a fuel and oxidizer (oxidizer denoted as LOX). As the fuel and LOX reach the main combustion chamber, they react with each other to produce thrust and move the spacecraft forward.

Most of the basic rocket engines are designed for propelling fuel out as fast as possible, either through combustion or pressure. However, each of these methods has its own unique limitations. For example, combustion temperatures can only reach a certain maximum before the engine starts burning itself, depending on the alloys used to build it. In a pressurized system, there is naturally a pressure limit for the tank —the higher the required pressure, the greater the reinforcements needed, which leads to heavier rockets, reducing their propulsion efficiency.

A. Current Technologies:

The problems mentioned for the pressurized and combustion rocket systems both lead to the engine delivering less material for the rocket to propel itself. The bulk of the literature review portion of this paper focuses on the technological advancements being explored for combustion rocket systems to overcome the current limitations in the field.

B. The Preburner:

One current method of increasing fuel output is to engineer a pump that transfers the fuel into the combustion chamber at a higher velocity.[15] However, this introduces a new problem that must be addressed—powering the pump. Having an extra battery would increase the weight of the rocket; instead of adding a battery, experts designed a way to reroute some of the fuel to power the pump with a turbine. This concept is known as a preburner, shown in Figure 2. [16]

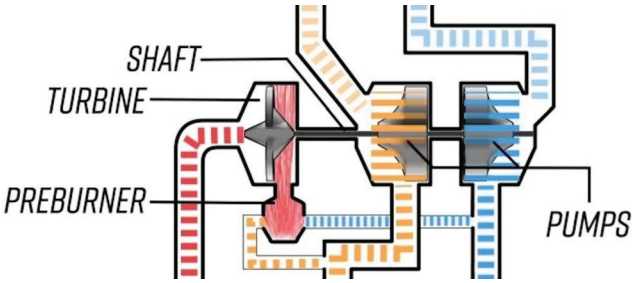


Figure 2: A standard preburner that helps spin the turbopumps.[10]

As a little bit of the fuel combusts in the preburner, it turns a turbine, which is connected by a shaft to the pump,[17] which throws all the fuel into the main combustion tank. But the extra burnt fuel doesn't just disappear. There are two main ways to deal with this exhaust—open-cycle and closed-cycle—and they are some of the key distinguishing factors in the four main rocket engines we will be discussing in this paper.

C. Open-Cycle:

The open-cycle engine is the most common gas generator bi-propellent engine that is running today.[18] This cycle engine is the most simplistic, as all the preburner exhaust is simply expelled overboard, as seen in Figure 3. The fuel that powers the preburner must have a higher pressure than the combustion chamber, as the combustion in the preburner turns the pumps

that supply the fuel and LOX into the combustion chamber, following the principle that pressure flows from high to low.

However, as pressure increases, so does temperature, which risks melting the entire engine system. In order to prevent this, as the preburner doesn't have the same cooling systems as the main combustion chamber, engineers run the preburner at a less stoichiometrically favorable ratio.[18] This causes the exhaust gas to become sooty, giving it a distinctly dirty appearance. This sooty residue can contaminate the combustion chamber, requiring engineers to carefully address this challenge in the preburner system. They decided the best design would be to dump this exhaust gas outside of the combustion chamber.

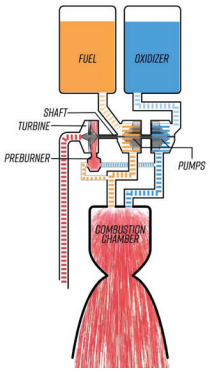


Figure 3: An open-cycle rocket engine that dumps the preburner fuel overboard on the side of the main combustion chamber. [10]

D. Closed-Cycle Engine:

A closed-cycle engine, unlike an open-cycle bipropellant engine, recirculates all preburner exhaust into the main chamber, a process that would damage an open-cycle engine due to its sooty, fuel-rich exhaust. So, engineers came up with 3 different solutions to solve this issue: the closed-cycle oxygen-rich (ORSC) engine, the closed-cycle fuel-rich engine, and the full-flow staged combustion cycle engine.

Closed Cycle Oxygen-Rich Engine:

In section II.C, it was stated that the preburner is run at a less stoichiometrically favorable ratio, which makes the preburner exhaust so sooty. So, engineers changed the design for all the oxidizers to be combusted in the preburner. However, this causes the temperature inside our preburners to reach up to 1383 K (~2030°F).[17] To prevent the engine from melting, engineers would have to develop an alloy that could withstand enormous amounts of heat. The United States couldn't find a way to solve this, as they believed that the alloy needed would be extremely difficult or even impossible to make. On the other hand, the Soviet Union was able to make this alloy: Inconel 718.[19] The alloy developed was a titanium alloy that could withstand temperatures of up to 973 K. By using this, the Soviet Union would be able to design the first closed-cycle oxygen-rich engine, shown in Figure 4.

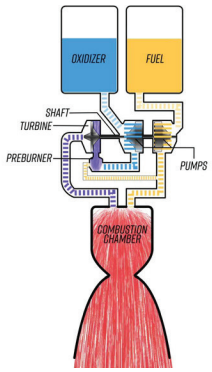


Figure 4: A closed-cycle oxygen-rich engine. Requires extremely strong alloys to prevent melting. [10]

Closed-Cycle Fuel-Rich Engine:

With the US not being able to make a strong enough alloy to withstand the temperatures of an ORSC engine, they opted to change the fuel and make a closed-cycle fuel-rich engine.

But as said in section II.C, the fuel-rich preburners typically produced a lot of "sooty" chemicals. However, this only happened when kerosene (RP-1), the most common fuel at the time, was used. So, the United States opted for a different fuel: hydrogen. However, just using hydrogen as a fuel wouldn't fix everything in the engine, and using hydrogen as a fuel came with its own set of problems. Hydrogen is a combustible gas, which is great for burning as a fuel, but not so great if the hydrogen hasn't even reached the engine yet. If there were any hydrogen leaks outside the main combustion chamber, the engine and the rocket would explode. Even worse, hydrogen is a hard gas to store, as the gas turns into a liquid at 20.1 K (−423°F). Additionally, the density of liquid hydrogen is 70.9 kg/m^3, lower than any other fuel at STP (Standard temperature and pressure).[20] There is not much scientists can do to solve this problem, and they can only mitigate it at best. As a result, huge insulated tanks are required in order to store the highly combustible fuel. Due to the difference in density between the LOX and hydrogen, two different preburners must be engineered—one is attached to the fuel pump, while the other is attached to the LOX pump.[24] In this system, all of the hydrogen gets pumped into the preburners, and that preburner's exhaust gas is pumped into the main combustion chamber. While some of the LOX goes into the preburners, the majority of it ends up combusting in the main chamber. This causes both preburners to run fuel-rich, creating the closed-cycle fuel-rich engine, as seen in Figure 5.

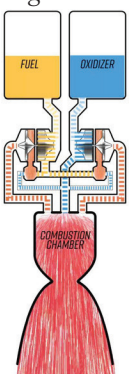


Figure 3: A closed-cycle fuel-rich engine. Both preburners run fuel-rich. [1]

DOI: 10.36838/v7i10.72

Full-Flow Staged Combustion Cycle Engine:

Given the critical role of engine design in optimizing engine efficiency, experts are always looking for ways to improve. In recent years, there has been a significant advancement in the development of one design, the full-flow staged combustion cycle engine, shown in Figure 6. The full-flow is similar to other staged combustion cycle engines in the sense that it is also a closed cycle engine and utilizes preburners. However, the specific way they use their fuel and preburners makes it unique. While the other combustion cycles have their preburners run either fuel-rich or oxygen-rich, the full-flow staged combustion cycle has two preburners, one running fuel-rich and the other running oxygen-rich.[21] The exhaust of both preburners enters the combustion chamber in a gaseous state and combusts. Because of this gaseous state, combustion becomes much more efficient. [18]

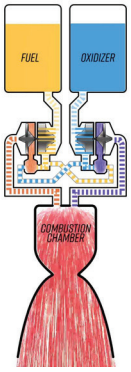


Figure 6: A full-flow staged combustion cycle engine. One preburner runs fuel-rich, the other preburner runs oxygen-rich. [10]

One factor that prevented the full-flow staged combustion cycle engines from being developed faster was the lack of a good fuel and a strong alloy.

As mentioned in II.C, with a fuel-rich preburner, if RP-1 is used, then the exhaust gas would be extremely sooty, clogging the engine. If we instead run it with hydrogen, then the density difference between the fuel and oxidizer would force the rocket to have two fuel-rich preburners. In the 20th century, a total of two of these rocket engines were made. One ran on hydrogen, the other on UDMH (unsymmetrical dimethylhydrazine). It wasn't until the 21st century that a third and much more efficient fuel would be used, methane.

As mentioned before, in order to run an engine with an oxidizer preburner, there needs to be alloys that can withstand high pressure and heat. When SpaceX started building their full-flow engine, they developed a special alloy called SX500 that is able to withstand 800 bar of hot gas, suitable for building this engine.

Since the development of the Raptor engines, much of the attention has been focused on the full-flow staged combustion cycle engine. As SpaceX continues to make rapid progress on its engine, it's apparent how using this design will significantly contribute to the efficiency of the rocket.

E: Future Technologies:

In terms of engine design, there are many more ideas that are slowly being tested or have been tested.

Note that the two most common types of rocket engines are mono-propellant and bi-propellant systems. Rocketdyne, the organization behind the development of the RS-25 and F-1 engines, once attempted to develop a tri-propellant engine.[22] After extensive testing, it was found that the engine achieved a specific impulse (ISP) of 542 seconds, the highest ever recorded. However, the engine's propellant composition, gaseous hydrogen, liquid fluorine, and liquid lithium, produced byproducts such as hydrofluoric acid, a highly toxic and corrosive substance, making the engine's use in practical applications extremely hazardous and essentially impossible.

Some other emerging ideas for rocket engine designs are plasma rockets, fusion propulsion, and nuclear propulsion.

As for fusion nuclear propulsion, the benefits are huge. They can generate large amounts of thrust for extended periods of time, and the fuel efficiency is unmatched. They could finish a Mars mission with five to ten times less propellant than chemical rockets. However, there are huge safety and scaling challenges.

Plasma rockets may revolutionize rocket engines, as they rely on electromagnetic fields rather than chemicals. Plasma rockets are extremely efficient with their fuel compared to chemical rockets, as they can send ions at speeds much higher than the speed of exhaust from a traditional rocket.[23] However, plasma rockets cannot generate the same level of thrust as chemical rockets. They can produce consistent thrust for long periods of time, so the fuel efficiency is extremely high, but the overall speed produced is lower.

■ Methods

III: Data analysis of propulsion parameters:

A: Introduction to key Equations:

Using the Python proptools package,[5] we observe the relationships between independent variables (Exit Pressure, Throat Area, and Chamber Temperature) and their impact on (Thrust, Mass Flow, and Specific Impulse) with the goal of understanding which relationships contribute to the optimal efficiency of a propulsion system. Though there is no one-size-fits-all, the analysis performed allows for a breakdown of the general components required for a successful rocket system.

There are many variables used in the calculations of thrust, mass flow, and specific impulse. For all calculations, chamber pressure, exhaust molar mass, and specific heat ratio were kept constant. There are specific equations that help determine different variables involved in calculating the dependent variables.

One variable is Cf, or the thrust coefficient of the rocket engine. Some variables used in this calculation were:

- P_e represents the exit pressure of the system.
- P_c represents the chamber pressure of the system. This is set as a constant of 100 bar in all simulations done.
- γ is the specific heat ratio of the exhaust gas. This is set as a constant of 1.2 in all simulations done.

It is calculated with the equation:

$$C_F = \left(\frac{2\gamma^2}{\gamma-1}\cdot\left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{\gamma-1}}\cdot\left(1-\left(\frac{p_e}{p_c}\right)^{\frac{\gamma-1}{\gamma}}\right)\right)^{\frac{1}{2}} \quad \text{(Eq. 1)}^{24}$$

Equation 1: This equation calculates the thrust coefficient of the engine. The thrust coefficient is unitless.

The thrust coefficient parameter, C_f, characterizes rocket nozzle performance to determine how well the nozzle expands the exhaust gas. The higher the C_f, the more efficient thrust production for a given chamber pressure and nozzle geometry, which is seen in output values of specific impulse and, understandably, thrust.

Another variable is C*, or characteristic velocity. Some variables used in this calculation were:

- T_c represents chamber temperature
- M_{molar} represents the exhaust molar mass

It is calculated with the equation:

$$c^* = \frac{\sqrt{\gamma\left(\frac{R_{univ}}{m_{molar}}\right)T_c}}{\gamma}\cdot\left(\frac{2}{\gamma+1}\right)^{-\frac{\gamma+1}{2(\gamma-1)}} \quad \text{(Eq. 2)}^{24}$$

Equation 2: This equation is used to calculate characteristic velocity. The units of characteristic velocity are meters per second.

The characteristic velocity, denoted as C*, is the effectiveness with which the combustion in a rocket engine produces high temperature and pressure. Another way to describe this variable is how efficiently a system converts propellant energy to thermal and kinetic energy. This variable is not dependent on nozzle geometry, but on gaseous properties.

Key parameters explored in the analysis described in Section III.B will be reviewed in Section III.A. Therefore, Section III.A provides the foundational relationships between key characteristics of an efficient (good to provide some kind of metric for efficiency) rocket propulsion system, specifically the impact of exit pressure, chamber temperature, and throat area on specific impulse, mass flow, and thrust.

B: Predicted/Expected Parameter Relationships:

This section presents the expected relationships between the key parameters discussed in Section III.A, using plots generated with proptools. Each sub-section showcases the impact on each parameter that is analyzed further in this project.

B1. Specific Impulse:

Specific impulse is a ratio of the thrust produced to the weight flow of the propellants. It is a measure of how effectively a rocket engine uses the fuel, with a higher specific impulse meaning the same amount of propellant will produce more thrust for a different engine. The most widely accepted equation for specific impulse is:

$$I_{sp} = \frac{F}{\dot{m}g_0} \quad \text{(Eq. 3)}^{24}$$

Equation 3: This equation calculates specific impulse based on thrust (F), mass flow ($\dot{m}$), and gravity (g_0).

which corresponds to the following equation from proptools:

$$I_{sp} = \frac{C_f c^*}{g} \quad \text{(Eq. 4)}^{24}$$

Equation 4: This equation calculates specific impulse using characteristic velocity (c^*), gravity (g), and thrust coefficient (C_f).

Thrust coefficient is unitless, characteristic velocity is meters per second, and gravity is meters per second squared. Therefore, the units cancel out to seconds, which is the unit of specific impulse. Here, plots are provided that showcase the behavior of specific impulse against various key parameters.

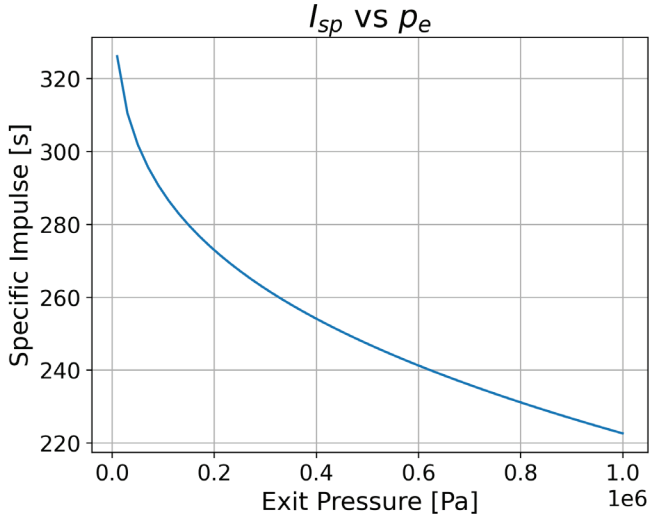


Figure 7: Specific impulse vs exit pressure. Plot depicting the expected behavior of specific impulse vs exit pressure for X system. As exit pressure increases, specific impulse decreases in a nonlinear manner, indicating diminishing performance losses.

From Equation 1, an increase in exit pressure will decrease the thrust coefficient and therefore decrease the specific impulse of the system. By graphing this relationship with *proptools* (Figure 7), we can confirm that as exit pressure increases, specific decreases by smaller and smaller amounts.

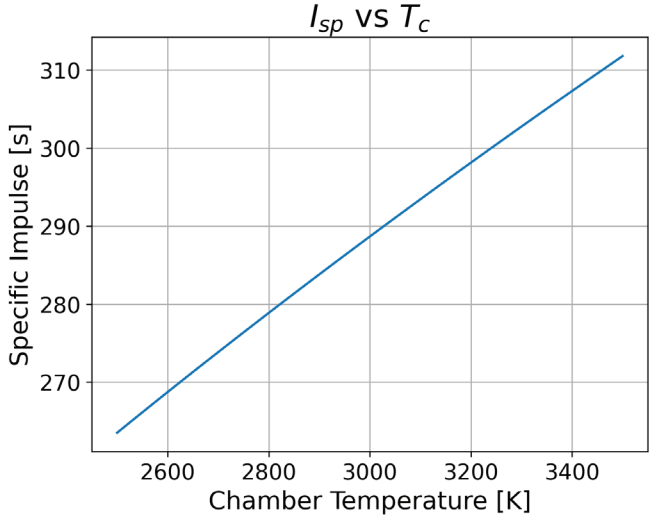


Figure 8: Specific impulse vs chamber temperature. Plot depicting the expected behavior of specific impulse vs chamber temperature for X system. Note that as chamber temperature increases, specific impulse increases by smaller and smaller amounts, reflecting diminishing gains in C*.

DOI: 10.36838/v7i10.72

From Equation 2, it is seen that chamber temperature has an effect on C*. As chamber temperature increases, C* also increases by smaller and smaller amounts, since C* is. By graphing this relationship on proptools (Figure 8), it is observed that as chamber temperature increases, C* does increase by smaller and smaller amounts, confirming the relationship found in the equations.

B.2 Mass Flow:

Mass flow is the rate at which a rocket engine uses up fuel to produce thrust. The more mass that an engine ejects, the more thrust the engine is able to generate.

The equation for mass flow is:

$$\dot{m} = \rho v a \quad \text{(Eq. 5)}^{24}$$

Equation 5: This equation calculates the mass flow based on fluid density (p), velocity (v), and cross-sectional area (a).

The equation used in *proptools* is:

$$\dot{m} = \frac{A_t p_c}{c^*} \quad \text{(Eq. 6)}^{24}$$

Equation 6: This equation calculates the mass flow based on throat area (A_t), chamber pressure (p_c), which is kept constant, and characteristic velocity.

The equation in proptools is confirmed through dimensional analysis. The units of mass flow are kilograms per second. The units of area, chamber pressure, and characteristic velocity are meters squared, newtons per meter squared, and meters per second, respectively. Substituting the units into the equation confirms that they simplify to kilograms per second.

Note: Exit Pressure is not a component for calculating mass flow, so it is not explored in this section.

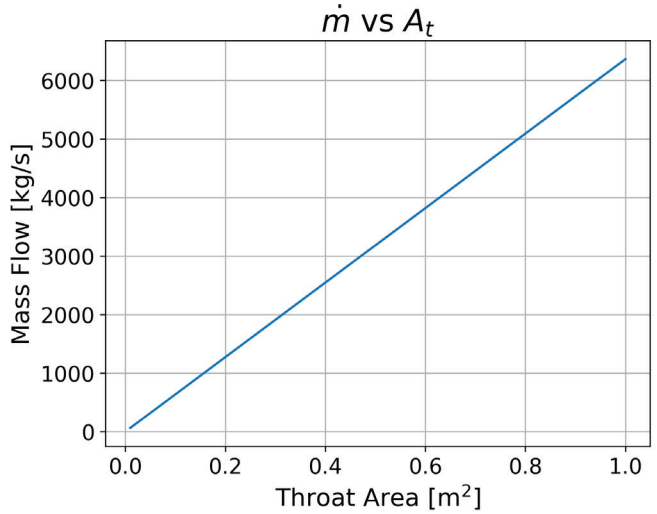


Figure 9: Mass Flow vs Throat Area. Plot depicting the expected behavior of mass flow vs throat area for increasing throat area values at a constant chamber pressure. Note that as the throat area increases, mass flow increases linearly, confirming the direct relationship predicted by Equation 4.

As shown in Equation 6, the throat area has a linear relationship with mass flow. By graphing this relationship on *proptools* (Figure 9), we can confirm that as throat area increases, mass flow increases by constant amounts.

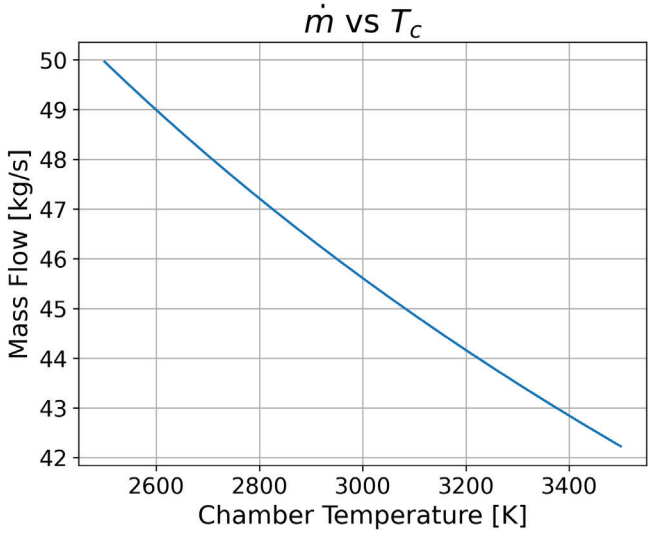


Figure 10: Mass Flow vs Chamber Temperature. Plot depicting the expected behavior of mass flow vs chamber temperature. Note that as chamber temperature increases, mass flow decreases at a decreasing rate, reflecting the inverse relationship through C*.

Chamber temperature is a directly related value in c_star, which is inversely related to mass flow. This means that as chamber temperature goes up, mass flow goes down. We can see this relationship in Figure 10.

B.3: Thrust:

Thrust is a force that propels an object forward. The more thrust a rocket can produce, the more work it can complete.

The commonly used thrust equation in rocketry is:

$$F = \dot{m} V_e + (p_e - p_0) A_e \quad \text{(Eq. 7)}^{24}$$

Equation 7: This equation calculates thrust with mass flow ($\dot{m}$), exhaust velocity (V_e), exit pressure (p_e), ambient pressure (p_0), and nozzle area (A_e).

The equation we used in proptools is:

$$F = A_t p_c C_f \quad \text{(Eq. 8)}^{24}$$

Equation 8: This equation calculates thrust with throat area (At), chamber pressure (Pc), and thrust coefficient.

We can confirm this equation by using dimensional analysis. The units of thrust are newtons. The units of area, chamber pressure, and thrust coefficient are meters squared, newtons per meter squared, and unitless, respectively.

Note: Chamber Temperature is not a component for calculating thrust, so it is not explored in this section.

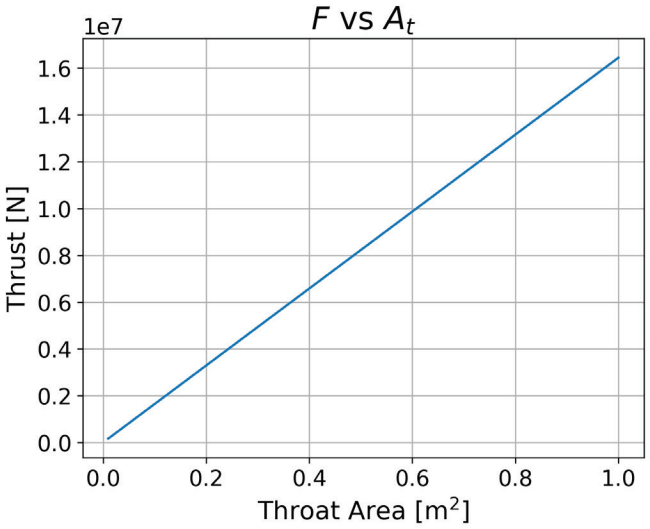


Figure 11: Thrust vs Throat Area. Plot depicting the expected behavior of thrust vs throat area for X system. Note that thrust increases linearly with throat area, confirming the direct relationship between the two variables.

The first independent variable analyzed in relation to thrust is throat area. Similar to mass flow, the throat area is linearly related to thrust, and therefore would produce a linear line, as shown in Figure 11.

The first independent variable analyzed in relation to thrust is throat area. Similar to mass flow, the throat area is linearly related to thrust, and therefore would produce a linear line, as shown in Figure 11.

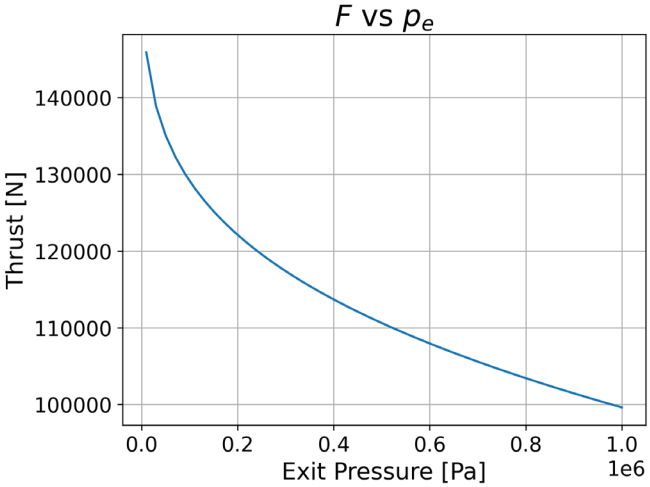


Figure 12: Thrust vs Exit Pressure. Plot depicting the expected behavior of thrust vs exit pressure for X system. Note that as exit pressure increases, thrust decreases nonlinearly, consistent with the reduction in thrust coefficient predicted by Equation 5.

The next independent variable analyzed is exit pressure. From Equation 1, it is seen that increasing exit pressure will decrease the thrust coefficient, which will decrease the thrust. We can confirm this relationship by plotting with *proptools* (Figure 12).

Result and Discussion

C: Influence of key parameters on gas and engine properties:

This next section demonstrates the effects of 2 independent variables on a dependent variable. The lines and curves produced in this analysis are due to gas properties/conditions and engineering components.

While some variables are dependent on the properties of the fuel, others are reliant on the engineering components of the engine.

Impact of Exit Pressure:

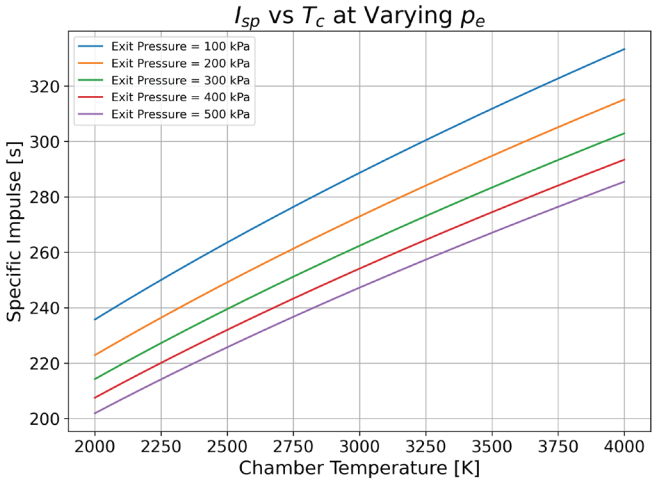


Figure 13: Relationship between Specific impulse and chamber temperature as a function of exit pressure. The relationship between specific impulse and chamber temperature is seen in this graph, matching the curve in Figure 8. As exit pressure increases, it can be seen that specific impulse decreases nonlinearly, as shown in Figure 7. This figure shows that specific impulse increases with chamber temperature but decreases nonlinearly with exit pressure, highlighting diminishing returns at higher pressures.

The overall shape of the Specific Impulse VS Chamber Temperature is in line with the graph in Figure 8. As seen in Figure 7, as exit pressure increases, specific impulse decreases by smaller amounts. However, the rate of this decrease becomes less steep with each additional increment of 100,000 Pascals, as illustrated in Figure 13. This indicates diminishing returns at higher exit pressures, while performance continues to drop, the marginal loss in specific impulse becomes smaller. This trend suggests that although higher exit pressures reduce engine efficiency, the impact lessens at elevated pressures, guiding tradeoffs in engine design and helping engineers determine how engines should operate under varying conditions.

DOI: 10.36838/v7i10.72

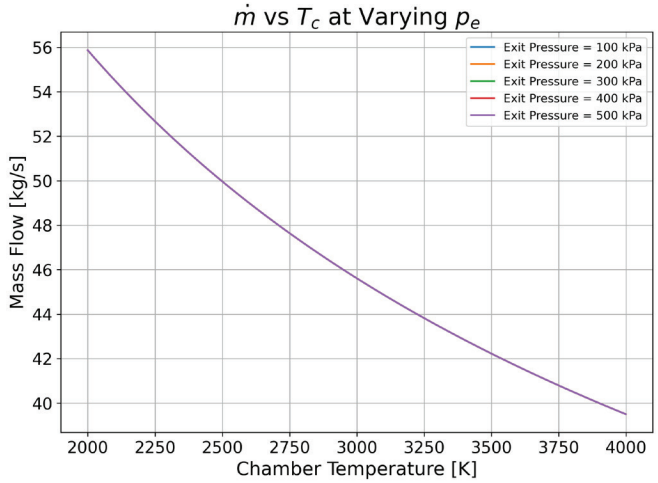


Figure 14: Relationship between Mass Flow and Chamber Temperature as a function of exit pressure. The shape of the curve is reflected in Figure 10. Since exit pressure doesn't affect mass flow, there is only one line shown on the graph. This figure shows that mass flow decreases with chamber temperature, matching the trend in Figure 10, and remains unaffected by changes in exit pressure, resulting in a single curve.

Since changing the exit pressure does not affect the mass flow of a rocket, as mentioned in section III.A.2, there is only one line displayed in Figure 14 for all the different exit pressures.

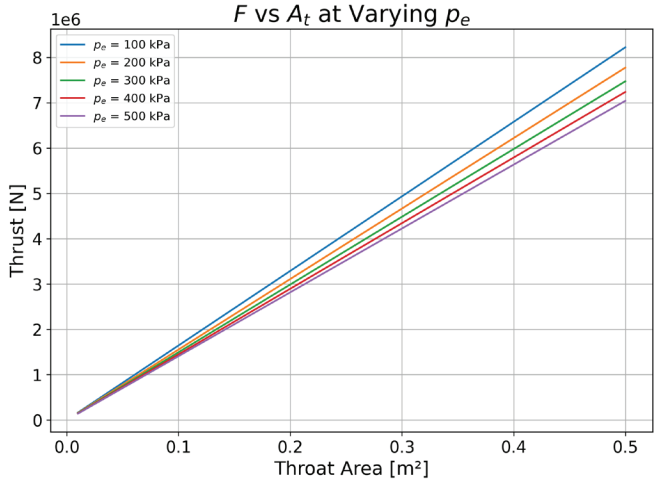


Figure 15: Relationship between Thrust and Throat Area as a function of exit pressure. The shape of the curve, a line, is the same as the one shown in Figure 11. Increasing the exit pressure decreases the thrust, a relationship demonstrated in Figure 12. This figure shows a linear increase in thrust with throat area, while rising exit pressure causes a nonlinear reduction in thrust.

The Relationship between Thrust and Throat Area is linear in this figure, preserving the expected shape seen in Figure 11. As exit pressure increases, thrust decreases by progressively smaller amounts, as seen in Figure 12. This is reflected in Figure 15, as the slope of the relationship between thrust and throat area decreases by smaller and smaller amounts. It is easiest to observe this decrease by looking at the points where the throat area is larger.

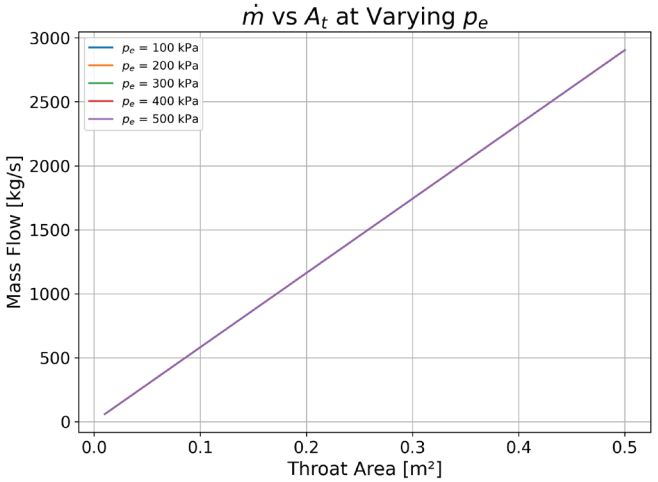


Figure 16: Relationship between Mass Flow and Throat Area as a function of exit pressure. The shape of the curve is preserved as seen in Figure 9. Since exit pressure doesn't affect mass flow, all the lines overlap. This figure shows mass flow increasing linearly with throat area, consistent with Figure 9, while exit pressure has no effect, resulting in overlapping lines.

Since exit pressure does not affect mass flow, as shown in section III.A.2, we observed only one line in Figure 16.

Impact of Chamber Temperature:

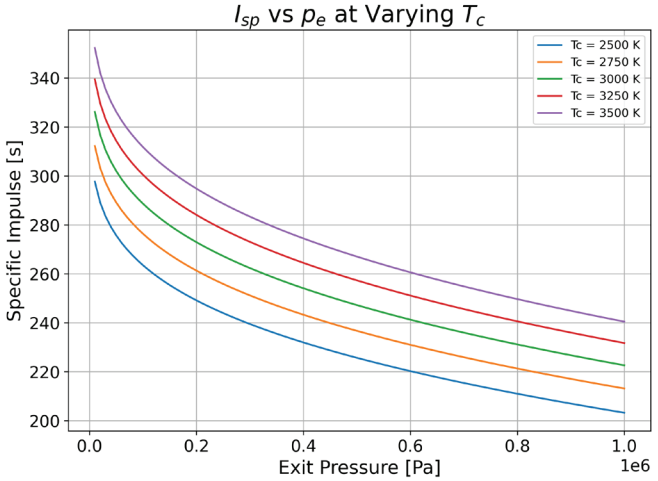


Figure 17: Relationship between specific impulse and exit pressure as a function of chamber temperature. Here, the shape of the curve between impulse and exit pressure remains as expected, as seen in Figure 7; however, as we increase chamber temperature, the amplitude of the curve changes, which implies that as we increase chamber temperature, we also increase specific impulse. This figure shows the expected nonlinear decrease of specific impulse with exit pressure, while higher chamber temperatures raise overall specific impulse values with diminishing incremental effects, suggesting a plateau at high temperatures.

The overall shape of the line representing the relationship between Specific Impulse and Exit Pressure in Figure 17 remains the same as seen in Figure 7. When plotting the relationship with different Chamber Temperature values, we can see that as Chamber Temperature increases at decreasing intervals, the gaps between the lines decrease. This is in line with the relationship we determined in Figure 7, where Specific Impulse decreases by smaller and smaller amounts as exit pressure increases at a constant amount. This also implies that there is a

theoretical limit to how much higher the Chamber Temperature can be before it has virtually no effect on Specific Impulse. This is shown in Figure 8, as Specific Impulse increases at decreasing amounts relative to Chamber Temperature, suggesting that the value of specific impulse will plateau.

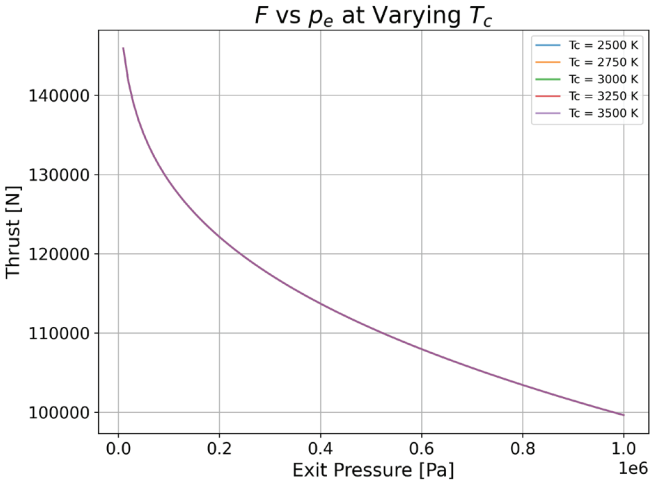


Figure 18: Relationship between Thrust and exit pressure as a function of chamber temperature. The shape of the curve between thrust and exit pressure is the same as seen in Figure 12. Changing chamber temperature does not affect thrust, as all lines with different chamber temperatures overlap with each other. This figure shows that thrust decreases nonlinearly with exit pressure, consistent with Figure 12, while chamber temperature has no effect, resulting in overlapping lines.

The legend in Figure 18 shows the chamber temperature values evaluated on the graph, and as seen, they do not affect the relationship between exit pressure in thrust. Since changing chamber temperature does not affect thrust, this graph has overplotted the relationship at various chamber temperatures on top of each other, emphasizing the lack of correlation between chamber temperature and thrust shows one line.

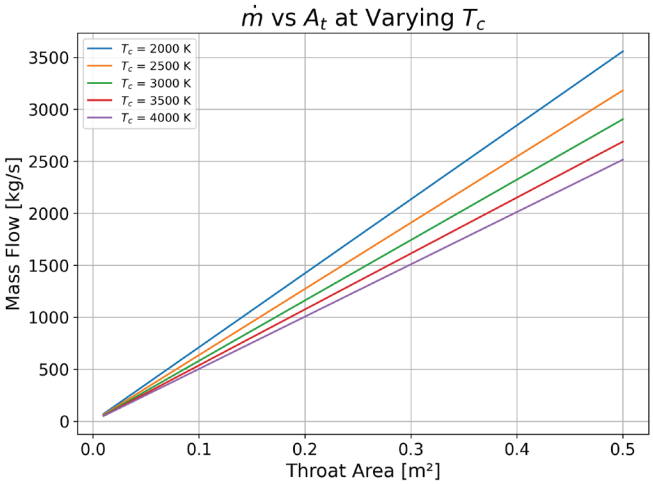


Figure 19: Relationship between Mass Flow and Throat Area as a function of chamber temperature. The shape of the curve for the relationship between mass flow and throat area is maintained as seen in Figure 9. As chamber temperature increases, mass flow decreases, as shown in Figure 10. This figure shows a linear increase in mass flow with throat area, while rising chamber temperature causes a gradual decrease in mass flow, reducing the slope as temperature increases.

As the throat area increases, mass flow also increases linearly, as shown in Figure 9. When the chamber temperature increases, mass flow decreases by smaller and smaller amounts. This leads to the slopes of the relationship between mass flow and throat area to decrease as chamber temperature increases, as seen in Figure 19.

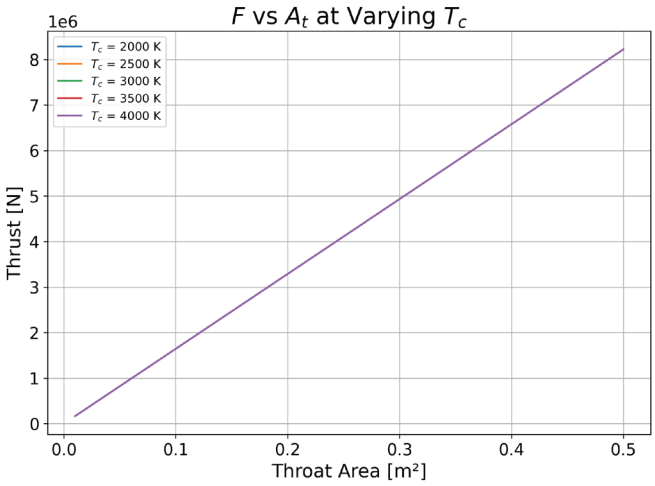


Figure 20: Relationship between Thrust and Throat Area as a function of chamber temperature. The shape is maintained from the simulation in Figure 6. Since chamber temperature doesn't affect thrust, all the lines plotted overlap. This figure shows a linear increase in thrust with throat area, unaffected by chamber temperature, resulting in overlapping lines.

Chamber Temperature does not affect thrust as seen in section III.A.3, so there will only be one observable line in Figure 20.

Impact of Throat Area:

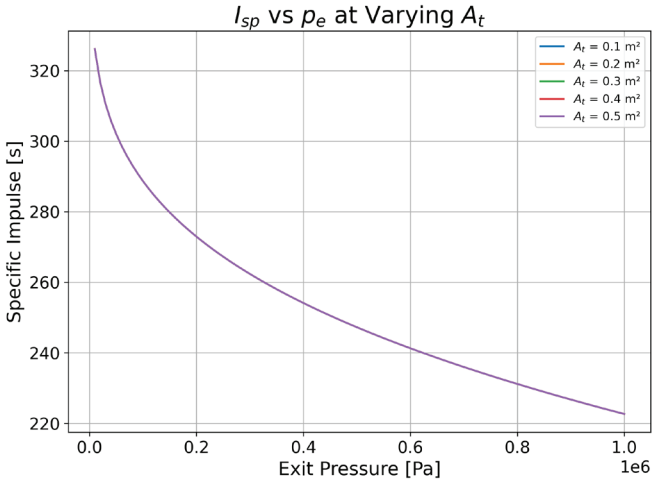


Figure 21: Relationship between Specific Impulse and exit pressure as a function of throat area. The shape of the curve is a reflection of Figure 7, which makes sense as it is plotting the same relationship. The throat area is determined not to affect the relationship, as all the lines overlap. This figure shows the nonlinear decrease of specific impulse with exit pressure, with throat area having no effect, resulting in overlapping lines.

The legend shows the throat areas tested in this simulation, and it is shown that all throat areas plot the same line. So, throat diameter also does not affect specific impulse, as seen in Figure 21.

DOI: 10.36838/v7i10.72

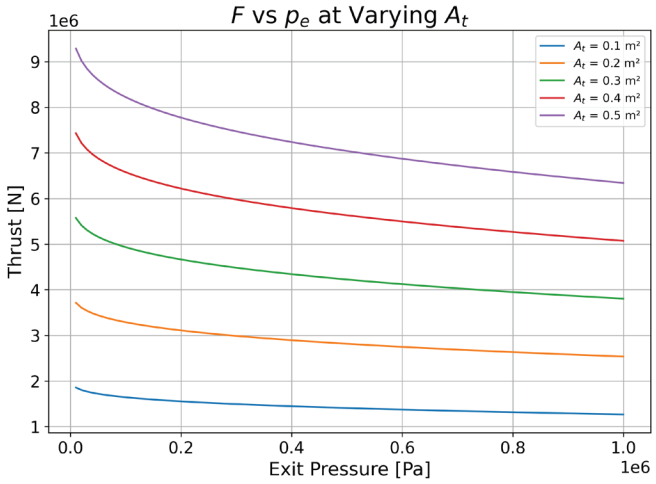


Figure 22: Relationship between Thrust and exit pressure as a function of throat area. The shape of the curve for the thrust vs exit pressure relationship is in line with Figure 12. As the throat area increases, it can be seen that thrust also increases, which is a relationship confirmed in Figure 11. This figure shows that thrust decreases nonlinearly with exit pressure and increases linearly with throat area.

The overall shape of the thrust vs exit pressure remains the same as seen in Figure 12. When factoring the throat area into the relationship between exit pressure and thrust, we can see that increasing the throat area increases the thrust of the system (plotted in Figure 22 as throat area). It seems that in order to increase the thrust, we should just increase the throat area. Recall that to generate this plot, chamber pressure was set to be a constant value, 100 bar. However, this is one key example of the edge cases with simulations/the proptools Python package; in a realistic setting, any change in throat area will impact chamber pressure, and therefore exit pressure. Recall that as the throat area increases, chamber pressure/exit pressure will decrease. Therefore, increasing the throat area while maintaining a high chamber pressure is difficult, making some of the thrust values in the graph unrealistic.

Note: Thrust is calculated with the thrust coefficient, which has chamber pressure as an input value.

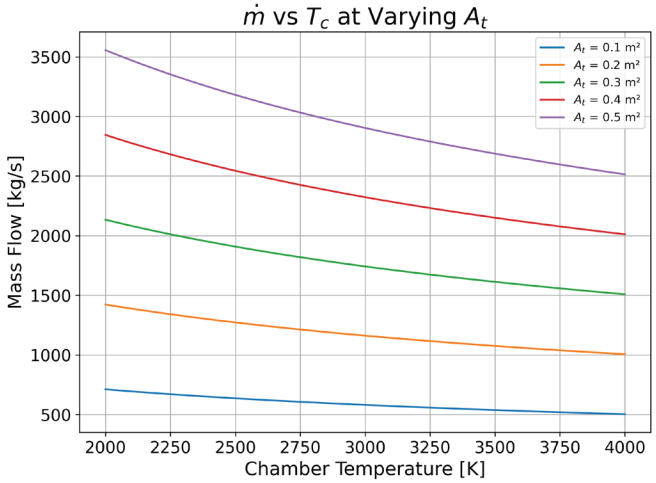


Figure 23: Relationship between Mass Flow and Chamber Temperature as a function of throat area. The shape is maintained from the simulation in Figure 10. As the throat area increases, mass flow increases, as shown in Figure 9. This figure shows mass flow decreasing with chamber temperature and increasing linearly with throat area.

The overall shape of the relationship between mass flow and chamber temperature is the same as in Figure 10. The relationship between mass flow and throat area is linear, as shown in Figure 9. The differences between the throat area cases are consistent, with uniform offsets separating the curves. Therefore, we can confirm that Figure 23 follows the same properties as Figures 9 and 10.

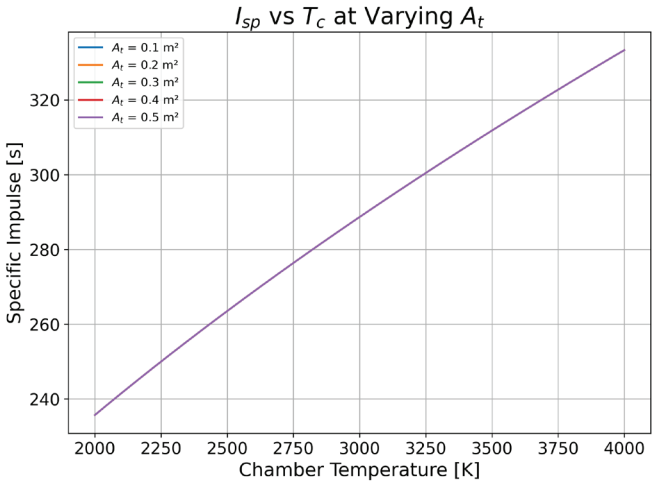


Figure 24: Relationship between Specific Impulse and Chamber Temperature as a function of throat area. The curve has the same shape as the simulation in Figure 7, which confirms the relationship between specific impulse and chamber temperature. Changing the throat area does not affect this relationship, as the throat area does not affect specific impulse, as mentioned in section III.A.1. This figure shows that specific impulse increases with chamber temperature while remaining unaffected by throat area, resulting in a single overlapping curve.

Since the throat area does not affect specific impulse, as shown in section III.A.1, this plot shows that any change in throat area will have no impact on the impulse. This may be realistically untrue as the throat area would have an impact on chamber temperature. Because the throat area is assumed not to have an impact, there is only one line for the relationship between specific impulse and chamber temperature (Figure 24).

Overall, the parameters are impacted by two of the independent variables observed. Specific impulse has a nonlinear decreasing relationship with exit pressure and a nonlinear increasing relationship with chamber temperature. Mass flow is impacted by throat area and chamber temperature. It has a linear increasing relationship with throat area and a nonlinear decreasing relationship with chamber temperature. Thrust has a linear increasing relationship with throat area and a non-linear decreasing relationship with exit pressure. These relationships result in the graphs seen in Section III.C, with some graphs displaying multiple lines and others displaying a single line. The errors observed in this analysis result from the variables kept constant. For all relationships, chamber pressure, exhaust molar mass, and γ were kept constant, even if they would have been impacted by the changing independent variable.

Conclusion

In conclusion, propulsion systems, specifically chemical ones, rely on the energy from propellants, converting this en-

ergy into thrust through fluid dynamics and thermodynamics. Factors such as high combustion chamber temperatures, high energy content propellants, small isentropic coefficients, and lower molecular weight in exhaust gases improve propulsion efficiency. The type of fuel used in bi-propellent engines also significantly affects performance, with RP-1, liquid hydrogen, and liquid methane being the most common options, each with its own advantages and disadvantages. Rocket engine designs, including open-cycle and closed-cycle engines, continue to evolve, with advancements in designs such as full-flow staged combustion cycle engines. Future propulsion technologies, such as plasma rockets, fusion propulsion, and nuclear propulsion, hold great potential, though they face challenges in scalability and safety.

This paper also analyzed how key propulsion parameters, specifically specific impulse, mass flow, and thrust, are influenced by variables like exit pressure, chamber temperature, and throat area. Using theoretical analysis and simulations with Python's Proptools library, we were able to examine the relationships between these factors and rocket engine performance.

Some conclusions made are that specific impulse is most affected by changes in chamber temperature and exit pressure. For example, at a fixed chamber temperature of 3000 K, specific impulse decreases from ~290 seconds at 100 kPa to ~250 seconds at 500 kPa, confirming the inverse relationship between exit pressure and specific impulse. However, this decrease is non-linear, and the rate of decrease diminishes with each 100 kPa increment, indicating diminishing returns at higher exit pressures.

On the other hand, exit pressure does not impact mass flow. Mass flow is directly proportional to throat area, increasing linearly as throat area expands. For instance, increasing the throat area from 0.1 m^2 to 0.5 m^2 raises mass flow from approximately 600 kg/s to 3000 kg/s, depending on chamber temperature.

Thrust also follows a similar linear trend with throat area, increasing significantly as throat area increases. However, unlike mass flow, thrust is influenced by exit pressure. At a chamber temperature of 3000 K, thrust decreases as exit pressure increases, dropping from about 1.3 MN to 1.1 MN between 100 kPa and 500 kPa, though again, the decline becomes less steep at higher pressures. This shows that while increasing throat area can improve thrust, optimizing exit pressure is crucial for maintaining or boosting performance.

Overall, this research provides a deeper understanding of how specific impulse, mass flow, and thrust interact in a rocket engine system. While the throat area plays a key role in mass flow and thrust, optimizing specific impulse requires more focus on exit pressure and chamber temperature. These findings are useful for future advancements in propulsion technology, as balancing these parameters will be important for building more efficient and effective rocket engines.

Acknowledgments

I would like to acknowledge Dr. Sanah Bhimani for mentoring me through the research process.

References

1. proptools: Rocket Propulsion Design Tools — proptools 0.0.0 documentation. Readthedocs.io. https://proptools.readthedocs.io/en/latest/ =
2. 4.4: Combustion. Heats of Reaction. Bond Energies. Chemistry LibreTexts. https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Basic_Principles_of_Organic_Chemistry_%28Roberts_and_Caserio%29/04%3A_Alkanes/4.04%3A_Combustion._Heats_of_Reaction._Bond_Energies.
3. Inglenook Engineering. Isentropic Expansion Coefficient Calculation. Inglenook Engineering. https://www.inglenookeng.com/isentropic-expansion-coefficient-calculation/
4. Ibrahim, T. K.; Rahman, M. M. Effects of Isentropic Efficiency and Enhancing Strategies on Gas Turbine Performance. JOURNAL OF MECHANICAL ENGINEERING AND SCIENCES 2013, 4, 383–396. https://doi.org/10.15282/jmes.4.2013.3.0036.
5. Ballard, R. Liquid Propulsion Systems - Evolution & Advancements Launch Vehicle Propulsion & Systems. https://ntrs.nasa.gov/api/citations/20140002716/downloads/20140002716.pdf.
6. Chaturvedi, S.; Dave, P. N. Solid Propellants: AP/HTPB Composite Propellants. Arabian Journal of Chemistry 2019, 12 (8), 2061–2068. https://doi.org/10.1016/j.arabjc.2014.12.033.
7. Halchak, J.; Cannon, J. Materials for Liquid Propulsion Systems. https://ntrs.nasa.gov/api/citations/20160008869/downloads/20160008869.pdf.
8. Helmenstine, A. Adiabatic Flame Temperature Chart. Science Notes and Projects. https://sciencenotes.org/adiabatic-flame-temperature-chart/#google_vignette
9. Perry, B. Liquid oxygen – Rocketology: NASA's Space Launch System. Nasa.gov. https://blogs.nasa.gov/Rocketology/tag/liquid-oxygen/.
10. Dodd, T. Is SpaceX's Raptor engine the king of rocket engines? Everyday Astronaut. https://everydayastronaut.com/raptor-engine/.
11. Marshall, W.; Kleinhenz, J. SUMMARY of ALTITUDE PULSE TESTING of a 100-LB F LO 2 /LCH 4 REACTION CONTROL ENGINE. https://ntrs.nasa.gov/api/citations/20120004055/downloads/20120004055.pdf
12. Mars Express confirms methane in the Martian atmosphere. www.esa.int. https://www.esa.int/Science_Exploration/Space_Science/Mars_Express/Mars_Express_confirms_methane_in_the_Martian_atmosphere.
13. NASA. Rocket Propulsion. Nasa.gov. https://www.grc.nasa.gov/www/k-12/airplane/rocket.html.
14. OpenStax. 9.7 Rocket Propulsion. pressbooks.online.ucf.edu 2016.
15. Hill, P.; Peterson, C. MECHANICS and THERMODYNAMICS of PROPULSION. https://soaneemrana.com/onewebmedia/MECHANICS%20AND%20THERMODYNAMICS1.pdf.
16. Force, A. HYDROCARBON BOOST (HCB) TECHNOLOGY DEVELOPMENT PROGRAM – Air Force Research Laboratory. Afresearchlab.com. https://afresearchlab.com/technology/successstories/hydrocarbon-boost-hcb-technology-development-program/.
17. Yost, M. C. PREBURNER OF STAGED COMBUSTION ROCKET ENGINE

DOI: 10.36838/v7i10.72

https://ntrs.nasa.gov/api/citations/19780016336/downloads/19780016336.pdf
18. Law, A. Rocket Engine Cycles. Everyday Astronaut. https://everydayastronaut.com/rocket-engine-cycles/.
19. INCONEL ® Alloy 718. https://www.specialmetals.com/documents/technical-bulletins/inconel/inconel-alloy-718.pdf.
20. DEMACO. The energy density of hydrogen: a unique property - Demaco Holland B.V. Demaco Cryogenics. https://demaco-cryogenics.com/blog/energy-density-of-hydrogen/.
21. Schoenman, L. FUEL/OXIDIZER-RICH HIGH-PRESSURE PREBURNERS. https://ntrs.nasa.gov/api/citations/19820002372/downloads/19820002372.pdf
22. Zurawski, R. Current Evaluation of the Tripropellant Concept; 2602. https://ntrs.nasa.gov/api/citations/19860018652/downloads/19860018652.pdf.
23. Crew, B. This Plasma Engine Could Get Humans to Mars on 100 Million Times Less Fuel. ScienceAlert. https://www.sciencealert.com/this-plasma-engine-could-get-humans-to-mars-on-100-million-times-less-fuel
24. Sutton, G. P.; Biblarz, O. Rocket Propulsion Elements; Wiley: Hoboken, N.J., 2010.

Author

Katherine is a junior at the Harker School in San Jose, California, with a passion for rocket science, control theory, and AI. She plans to pursue a career in the aerospace engineering or computer science industry.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

The Housing Crisis in U.S. Mountain Towns

Matthew Hagedorn

Aragon High School, 900 Alameda de las Pulgas, San Mateo, CA, 94402, USA; matthagedorn07@gmail.com

ABSTRACT: There are more than a hundred mountain towns across the United States, each with its own unique town culture, history, and natural amenities. Still, they share distinct economic dynamics like seasonal revenues and reliance on the tourism industry. Tourism drives business revenues, job opportunities, and overall economic prosperity, but it also has its drawbacks, especially when demand outstrips supply. A particular challenge is the housing shortage and rising prices, which threaten the accessibility and affordability of mountain towns. The development is primarily driven by the towns' growing popularity and a housing supply constricted by environmental, financial, and regulatory constraints, pressuring prices upward. The housing development threatens affordability and the displacement of long-standing community members, essential workers like teachers, firefighters, and medical workers, and the service workers who support the tourism industry that attracts economic revenues in the first place. These potential consequences highlight the delicate balance between economic growth and long-term sustainability in mountain towns. In response to the affordable housing concerns, many mountain towns have adopted or emphasized policies across the spectrum of command-and-control to market-based solutions. Examples of such public policies include more lenient regulations, deed restrictions, and financial incentives through tax credits, rebates, or direct subsidies.

KEYWORDS: Behavioral and Social Sciences, Sociology and Social Psychology, Tourism Economics, Affordable Housing, Sustainable Development.

Introduction

Mountain towns are distinctive landmarks to the U.S.'s terrain; nestled within various mountain ranges like the Appalachians, Sierra Nevada, and Rocky Mountains, mountain towns offer scenic landscapes, natural beauty, adventurous opportunities, and often culturally charming towns with deep history.[1] Their beauty does not go unnoticed, and yearly, mountain towns in the United States attract millions of visitors through their enticing outdoor recreation activities; the United Nations Environmental Programme estimates mountain towns make up about 15-20 percent of all tourism worldwide.[2] In 2023 alone, outdoor recreation generated $1.2 trillion in economic revenue and provided five million jobs.[3] These towns play a vital role in the U.S. economy, providing memorable and exciting experiences for many and driving the tourism industry, which supports local businesses, service jobs, and brings life to many longstanding and charming communities. Yet, all great things come with a cost, and beneath the idyllic exteriors lie economic and social complexities that threaten the towns' long-term sustainability and accessibility.

The cultural and economic importance of mountain towns extends beyond their tourism appeal. Their long-term success depends on maintaining a balance between fostering economic growth, supporting tourism, preserving attractive environmental quality, and protecting the needs of existing communities. The local economies depend greatly on their natural amenities and service workers. A prime example of this reliance is the popular skiing and snowboarding industry, which hinges greatly on ideal climate and environmental preservation. In the 2022-2023 season, ski resorts experienced a record 65.4 million visitors.[4] However, unchecked growth in these areas presents a tradeoff. As the popularity of outdoor recreation and mountain towns increases, more visitors can strain local infrastructure, cause housing affordability to fall, and even promote environmental degradation.[5] Without proper management and vigilance, these potential consequences threaten a prosperous future for mountain towns.

This paper is a policy review, examining the intersection between environmental economics in U.S. mountain towns and the challenges they face. It begins by defining mountain towns based on geographic, environmental, and economic criteria, particularly the prominent tourism industry. The paper then delves into the benefits and drawbacks of a tourism-centric economy, focusing on the tension between growth and social equity. The primary focus of this paper is the challenge of housing affordability in mountain towns and the various policies intended to address the negative impact on local workers and the long-term accessibility of mountain towns. Policy analysis covers market-based, command-and-control, and hybrid policies, exploring how mountain towns approach a balance between economic growth and social goals. By presenting this information and analysis, this paper intends to offer readers a detailed understanding of the importance of mountain towns, the challenges they face, and current policies to guide constructive steps forward.

Result and Discussion

Defining Mountain Towns in the United States:

Mountain towns are renowned for their natural wonders and seeming escape from urban society and bustling cities. They are often romanticized for providing a quieter lifestyle marked by awe-inspiring scenery, abundant wildlife, and outdoor rec-

DOI: 10.36838/v7i10.84

reation across grand mountains, through deep forests, and by pristine rivers and lakes. Mountain towns are hubs of environmental and cultural elements that cultivate their popularity among many. No two mountain towns are the same; each provides diverse natural landscapes, opportunities, and beauty.[6] However, mountain towns do share some similarities across the United States, and this paper will seek to define mountain towns in the United States by drawing parallels between their geographical, population, and economic qualities.

Geographic elements define mountain towns in the United States. The most obvious is that a mountain town must be situated near or within a mountain or mountainous region.[7] To qualify as a mountain, the United Nations Environment Programme and its work with the World Conservation Monitoring Centre use elevation to classify mountain-town regions between 300 and 2500 meters. To account for some highland plateaus or tablelands, mountain towns are further defined as having access to mountains and their region marked by steep slopes and a high variance in elevation within a small area.[8] Mountain towns are also identified by being near natural amenities that provide both visual contentment and the possibility for outdoor recreation. Amenities like rivers, lakes, bodies of water, forests, mountains, and wildlife are primary drivers for mountain towns' popularity.[9]

Many U.S. mountain towns experience unique population dynamics. Research done in Colorado's major mountain towns suggests the populations of mountain towns are made up of permanent residents, second homeowners and part-time residents, and tourists.[10] These observations are not isolated to Colorado's towns but apply to mountain towns all across the United States. Recent changes in population and economic dynamics are largely attributed to the COVID-19 pandemic accelerating the adoption of remote working for many and shifting preferences toward second homes in mountain towns. In Colorado's Routt, Grand, Eagle, Summit, and Pitkin counties, a 2021 study reported that about 50% of households in these mountain towns have at least one person working from home; that statistic rose to 58% in a 2023 survey.[11] Although the number has dropped since the peak of the pandemic, 37% of Colorado residents report a remote working schedule for at least one day a week.[12] Consequently, second homeowners, part-time residents, and visitors have become a growing population in mountain towns, exacerbating the influx of visitors experiencing mountain towns. Mountain towns face distinct seasonality, driving visitor trends; colder seasons attract visitors for winter sports like skiing and snowboarding, while warmer seasons attract hikers and mountain bikers. The population dynamics in mountain towns are uncommon and changing, but they also strongly influence the economic activity, trends, and culture in mountain towns.

A notable attribute of mountain towns is their distinct economic drivers and activities. Economic growth in mountain towns is spearheaded by tourism and other part-time visitors, most of whom participate in the outdoor recreation industry, which contributes to economic revenue and supports jobs.[13] The growing number of visitors in recent years has brought unprecedented growth and development to mountain towns, generating a large proportion of their revenue through outdoor recreation and related activities.[14] However, economic growth does not come without growing pains, and in many towns, more visitors result in overstressed public infrastructure, environmental degradation, rising prices, and public frustration. A leading concern is the skyrocketing housing and living prices that are making the lives of the local workforce unsustainable. The duality between rapid development and a sustainable way of life prompts a need for review. This paper intends to unpack the economics of mountain towns to assess current and potential threats and evaluate and propose constructive steps forward.

Social and population dynamics work together closely to produce economic development in mountain towns. A recent development is the introduction of more visitors and residents from urban areas and cities, who have jobs outside the mountain regions and are of a higher socio-economic class than the local workforce. The introduction of wealthier people has paradoxical effects. Their growing presence fuels economic growth but also brings various developmental challenges, as these new visitors have higher incomes than most locals.

The General Economics and Challenges of Mountain Towns:

The economics of mountain towns are complex and concern geographical, social, and environmental constraints. Notably, tourism and related activities surrounding outdoor recreation are the chief economic engines for mountain towns and their development. Tourism has many drivers, like family, business, or culture, but economically, tourism is the activity of individuals traveling to a different location for less than a year, bringing economic, environmental, and social consequences by influencing the supply and demand for goods and services, subsequently allocating resources toward specific needs.[15] The United Nations World Tourism Organization furthers this definition by classifying "mountain tourism" as a unique form of tourism existing specifically within mountainous areas with distinct climates, biodiversity, and natural amenities, as defined in earlier paragraphs.[16] For mountain towns, fast-growing popularity draws in more visitors from non-mountain-town native areas, affecting the economic and social dynamics within mountain towns, ultimately exposing many towns to unique and burdensome problems as infrastructure is stressed, community priorities are challenged, and the trajectory of the towns' futures evolves.[17]

Tourism growth in mountain towns is because of the unique natural amenities available. Mountain towns and especially those near national and state parks act as "gateway communities" for tourists to connect with attractive landscapes.[17] Preferences for the outdoors, nature, and recreation pair well with mountain towns' provision of clean and fresh air, water, varied terrain, scenic views, and natural resources to foster exciting outdoor adventure. Popular recreation that brings economic activity includes hiking, sightseeing, swimming, visiting national parks, water sports, visiting and touring cultural locations, snowboarding, skiing, photography, wildlife watching, mountaineering, and more.[18] The challenging balance comes from tourism and the environment's strong interrela-

tion, representing a delicate balance that generates tremendous economic revenue for mountain towns.

The tourism industry generates tremendous economic revenue for mountain towns, especially those in mountain states like Montana, Colorado, Wyoming, and Utah. In Montana, from 2020 to 2021, the outdoor recreation economy grew by almost 30%, and in 2021 alone, real gross output for the outdoor recreation economy grew by 22%, providing a rise in employment by 13% and compensation by 16% in the industry.[19] In Colorado's Estes Park, a 2021 report recorded $3,270 in local tax receipts per resident household generated by travel-related spending. The travel-related economic activity generated 3,100 jobs in Estes Park alone, and in the entire state of Colorado, the travel industry grew by 42% in 2020.[20] Wyoming and its mountain towns, it is no different. In 2023, travel spending totaled $4.8 billion, a 7% increase from the previous year, and 33,470 jobs were a result of travel spending.[21] In the popular mountain state of Utah, tourism and visitors spent a record $12 billion in 2022, directly generating 98,600 jobs and indirectly supporting another 53,200 jobs. The majority of visitor spending in 2022 was on lodging, transportation, and dining out expenses, emphasizing the importance of local infrastructure and businesses to support increased tourism.[22] From a broader perspective, tourism and travel-related economic activity are instrumental to the development of mountain towns.

Winter sports and recreation are a critical component in tourism revenue for mountain towns, as demonstrated in states like Colorado, Utah, and North Carolina. An economic impact report from the National Ski Areas Association on the 2023-2024 season reveals that across the United States, downhill snow sports generated $58.9 billion in economic revenue and supported 533,000 snowsport jobs.[23] The growing popularity in the industry supported wage growth of 49% over the past five seasons and has attracted an investment of $757 million to set.[23] A 2015 study revealed that Colorado's downhill sports industry generated $4.8 billion in economic activity, directly supporting over 46,000 jobs in recreation, retail, lodging, and services, and the industry has grown since.[24] In Utah, one of the most popular states for skiing, snowboarding, and winter recreation, a phenomenal 2022-2023 season brought $2.6 billion in skier spending, representing almost a 32% increase compared to pre-pandemic rates in 2019-2022, highlighting the economic transformation the Covid-19 pandemic contributed to for U.S. mountain regions.[22] In North Carolina, home of popular resorts like Sugar Mountain Resort, Wolf Ridge, and Cataloochee, the total economic impact from direct spending for the state's ski industry in the 2022-2023 season was $148 million, but after considering both direct and indirect spending, the total economic value amounted to $244 million.[25] Beyond aggregate value, the 2022-2023 winter season in North Carolina fostered 1,760 jobs and more than $16 million in capital expenditures for ski resorts in the region, marking successful reinvestment into the industry.[25] Across the United States' mountainous regions, mountain towns prosper from the ski industry and the economic activity it attracts.

Increased tourism and subsequent economic development often come at the cost of negative environmental externalities. Additional man-made structures are constructed to sustain tourist activities, threatening the mountains' many unique ecosystems and habitats, plants, and animals. Infrastructure like roads, tracks, pathways, slopes, and other clearings used for transport, travel, or recreation incur the cost of habitat destruction, soil erosion, deforestation, and potentially water pollution. In response to surging demand, construction for new housing is necessary but can come at a high cost depending on its execution. New housing construction directly alters existing ecosystems through deforestation, soil erosion, and other alterations. In theory, constructing more housing will ameliorate the current housing shortage burdening many U.S. mountain towns, but may also enable further overcrowding of roads, public spaces, and other infrastructure by increasing the residential population size.[26]

Another environmental concern is that business people and firms in the tourism industry can damage the environment if they are hyper-focused on generating profit, and the market misrepresents all relevant factors. In mountain towns, hyper-competitive and profit-oriented behavior will bring economic growth, but such behavior frequently overlooks environmental costs.[27]

Empirical research in Tanzania, a country where tourism contributed 10.7% to GDP and 11.1% to national employment in 2020, illustrates this connection.[28] As tourism grows, so does the demand for housing, through hotels, vacation rentals, and resorts, along with transportation, energy, and waste services. Without the proper policies to regulate economic development, expansion leads to deforestation, habitat destruction and displacement, ecosystem disruption, pollution, and numerous other drivers of environmental degradation. The study in Tanzania provides empirical evidence suggesting that increasing tourism revenue leads to higher ecological footprints and larger energy consumption, particularly when development relies on fossil fuels.[28]

These outcomes stem from a traditional market failure: environmental externalities. Firms and individuals hyper-focused on profit often capitalize on the opportunities from climbing tourism without addressing costs not explicitly reflected in the price system. Costs not fully represented in the price system are pollution, land degradation, and biodiversity loss. These costs affect the broader community, especially in the future, but are not fully reflected in the market.[29] Because the market fails to price all the costs into such actions, firms and individuals may over- or under-allocate resources toward such actions, leading to exaggerated consequences. A relevant example of an environmental externality is air pollution from increased production and energy usage. Without regulatory foundations in place to force actors in the market to internalize these costs, firms and individuals may not directly assume the long-term environmental cost of air pollution, and thus may overproduce and overconsume. When all the factors in a decision are not reflected in the market, decisions become less efficient.[30]

Although tourism benefits mountain towns in terms of economic activity and growth, the nature of tourism creates seasonal economies that stress mountain towns and their local businesses and service workers. The influx of visitors prevalent

DOI: 10.36838/v7i10.84

in peak winter and summer seasons causes problems like overcrowding, economic inequality between wealthier visitors and locals, increased prices, and stressed infrastructure.[31] From the perspective of local businesses, inconsistent demand challenges year-round sustainability for local businesses as they struggle to navigate inconsistent consumer demand and a seasonal local workforce. For workers, seasonal employment exaggerates the generally lower wages service workers are paid, creating quality-of-life concerns for local workers.[32,33]

Conditions for local workers have worsened under the current housing shortage and rising prices, a crisis predominantly caused by the increase in remote workers whose higher-income jobs allow them to bid up prices beyond what is reasonably affordable for a service worker.[34] For example, in Colorado, housing prices have risen considerably across counties, notably in Pitkin County the median house price at the start of 2019 was $5.2 million, compared to the start of 2024 where the median home price was $11.5 million, representing over a 100% increase in median home price.[35,36] In Colorado's Grand County, the median home price changed from $568,443 to $1.2 million between the start of 2019 and 2024, representing another over 100% median home price increase.[33,34] Lake Tahoe, situated in California and Nevada, has experienced similar price spikes. In North Lake Tahoe, the median price for a home is $1.1 million, a 129% increase from before the pandemic. In nearby Truckee, median home prices were at $1.1M in the fall of 2021, a 44% jump from the start of the pandemic.[39] A report from Full Stack Economics displays a similar change. Prominent U.S. mountain towns like Salt Lake City and Ogden in Utah have both experienced an average home price increase of over 50% since 2017, outpacing the national average of a 33% increase.[40] Although rising housing prices are a nationwide threat to the United States, the distinct rise of housing prices in U.S. mountain towns threatens the towns' economic sustainability by making housing for essential service workers, teachers, hospital staff, and infrastructure workers unaffordable.

Critical Challenges in Housing:

The interplay of supply and demand in mountain town housing markets has created a housing crisis for mountain towns across the United States in recent decades. On the demand side, growing popularity in mountain towns has driven up home and rental prices. Such price changes are exacerbated by the COVID-19 pandemic, which introduced more affluent buyers looking for permanent residency or a second home to the market. These higher-income buyers further shift demand and inflate the costs of all home prices.[41] On the supply side, geographical, environmental, and economic limitations make it difficult for the housing market to correct itself. Many mountain towns are restricted by forests, rivers, mountains, hills, and the large percentage of land that is publicly owned, making buildable land scarce. Coupled with concerns for environmental protection, building new homes becomes a difficult and expensive problem.[41] In recent years, the economic conditions in mountain towns have stressed the housing market. Shifting preferences favoring mountain towns and greater incomes among buyers have shifted demand, inflating mountain town housing prices. On the other hand, challenging and expensive conditions for housing construction have limited supply, keeping up with the higher demand, allowing prices to be more easily bid up.[10]

Colorado and its many mountain towns exemplify the housing market and its supply and demand challenges. Before the Great Recession, Colorado had a surplus of housing units relative to households, but since then, the rapid growth in households has outpaced the minting of new housing units. Paradoxically, the housing crisis that is challenging Colorado's mountain towns is a result of Colorado's success. The beautiful nature, inviting towns, and available opportunities encourage more residents and visitors.[42] New visitors lead to new buyers, and despite the previous surplus in housing, a 2022 report suggests 325,000 new homes would need to be added over the next few years to restore housing stability to its historical rates.[43]

Although higher housing prices cause challenges for everyone, housing unaffordability is the biggest problem for lower-income households making less than $50,000. On average, these households spend 35% of their incomes on housing, a problem deemed financially burdening and problematic by the National Low Income Housing Coalition. The NLIHC claims that if more than 30% of a household's income is spent on housing, it could impede the possibility of other essentials like food, utilities, insurance, and childcare.[44] In total, the Colorado households that earn less than $50,000 spend an additional $2 billion supporting their housing by paying over the 30% standard. This additional $2 billion represents the potential for a huge economic stimulus for local businesses if housing were affordable enough for lower-income households.[45]

The housing shortage in Colorado is a serious threat to residents' quality of life, as indicated by a Centennial State Survey conducted by Colorado Mesa University. The 2017 survey recorded housing affordability and availability as the number one problem facing their communities, with 63% of respondents arguing there is insufficient availability of affordable housing.[42] Also, the proposed 325,000 new homes needed to establish housing stability in Colorado are not only unlikely but would come with their consequences. The construction of 325,000 new homes would come at the cost of increased population density, potentially straining infrastructure and natural resources, and damaging the dynamic of mountain towns through overcrowding.

When addressing the market, the popular question is: "Why has the increased demand not brought about a market correction by driving up supply?" The answer is complex and encompasses resource availability and costs, short-term rentals, the aftermath of the Great Recession, and investor behavior. Building costs and available land to build discourage supply by making construction often inaccessible and expensive. The popularity of short-term rentals in Colorado's mountain towns exacerbates the issue by introducing more demand into the market and encouraging rental properties over properties for residents.[46] The Great Recession was another contributor to the problem because many real estate firms faced bankruptcies and consolidations, reducing the number of suppliers in the

housing market. Many investors in the market have also cut the market's supply by buying up single-family homes, further reducing the availability of residential homes for people living in mountain towns for longer periods. The interplay between fewer suppliers and a competitive market fosters an environment where options for long-term residency are scarce and expensive.[45]

Montana, home to popular mountain towns like Whitefish, Bozeman, Big Sky, and West Yellowstone, has experienced transformative home prices in recent years. Notably, from 2019 to 2022, Montana homes increased in value by 33% when accounting for inflation. These changes are a result of shifting population dynamics and their effects on Montana's economy. The catalyst of these changes was between 2020 and 2021, when Montana's population increased by 28,000 more people than expected, a net migration rate 3.3x faster than the pre-pandemic average. Previously, this paper has discussed how the pandemic introduced more affluent people to mountain town areas, and Montana is no different. In the 2020-2021 migration surge, 45% of migrants older than 20 had a bachelor's degree, a statistic 37% greater than the non-migrant share. These new migrants bring higher education and higher incomes from their remote jobs, providing many of the new migrants with the resources necessary to buy homes and bid up prices. The evidence supports this behavior, as historically, 30% of migrating households to Montana owned their home within a year; however, since the pandemic migration, this value has risen to 45%. The population and economic changes in Montana serve as an opportunity to improve the economic state of Montana; however, the changes also threaten to exacerbate scarcity and the burden of infrastructure and the housing market.[47]

The mountain towns in Montana, Lake Tahoe, and Truckee are other examples of mountains experiencing economic shifts due to new migrants. Many San Francisco and Bay Area workers brought their higher salaries to the nearby Lake Tahoe, bringing higher housing prices and rents with them. Like other mountain towns, Lake Tahoe's housing is made up of many second homes and properties reserved for short-term rentals. A 2020 article by the San Francisco Chronicle reports that 54% of the 13,000 homes in North Lake Tahoe and Truckee are second homes, and 13% of those homes are short-term vacation rentals. The combination of limited housing stock and more demand from more affluent households is threatening the sustainability of Lake Tahoe's economy, as many service workers report having a difficult time securing housing and, as a result, are considering leaving the mountain town.[48]

Accessibility is a leading concern for mountain towns' futures, so naturally, the restricted housing supply is commonly considered the primary driver of increasingly unaffordable housing prices. In response to this belief, the neoclassical market-based housing supply (MBHS) theory suggests that a deregulation of development and land-use constraints would increase affordability by addressing the shortage.[49] However, Steffen Wetzstein's 2021 research suggests the reality is more complex than a simple supply-and-demand problem. He argues that underlying factors and inequalities will prevent simple deregulation from reducing housing prices for all income levels, as deregulation will lead mostly to more high-end housing developments.[49] Thus, policies to improve housing accessibility must address the total housing stock and directly support affordable housing units for lower-income individuals and essential workers. Fortunately, many mountain towns are taking hybrid approaches that blend market efficiency with policies targeted toward supporting those in greater economic need.

Public Policies Addressing Affordable Housing:

Mountain towns in the United States face significant challenges in housing, workforce sustainability, environmental impact, and overall economic development. To address these issues, a range of policies and programs have been implemented across the United States' mountain states. These policies and programs fall into two distinct categories: market-based policies and command-and-control policies.

Market-based policies and command-and-control regulatory policies represent the two major economic policy categories to manage mountain town challenges. Market-based policies intend to encourage specific behaviors while maintaining the market aspect of decision-making, where individual consumers and producers can determine how to allocate their resources. A common use of market-based policies is to address and incorporate the externalities, or external costs or benefits of production or consumption, to adjust the market to more accurately reflect the true costs and benefits of goods or services. This type of solution is economically advantageous because it aligns with the ideas of economic efficiency and maintains the flexibility of a free market.[50] However, these policies can be more prone to public resistance and backlash relative to command-and-control policies because the visibility of costs, such as environmental taxes, can be frustrating, especially for those who distrust the government or perceive the policies as profit-driven. In contrast, command-and-control policies impose direct regulations or standards, such as limiting certain developments or requiring specific practices. While often criticized for being potentially less economically efficient, the absence of direct and visible cost increases or revenue collection can result in command-and-control policies being perceived as more favorable or less intrusive and financially burdensome by people.[51]

Market-Based Policies in Mountain Towns:

Market-based policies are key tools for addressing housing affordability, and they do so by leveraging economic incentives while preserving market mechanisms to encourage the private sector to provide more affordable housing. These policies often work through financial incentives like tax credits, housing subsidies, and other funding supported typically by tax revenue.[52] In mountain towns where land restrictions and tourism-driven demand make affordable housing scarce, market-based policies can offer flexible solutions to encourage desirable outcomes.

Tax incentives through tax credits and rebates are an example of a market-based policy often used to encourage specific behaviors from private developers and individuals in address-

DOI: 10.36838/v7i10.84

ing housing affordability and availability. The policy tool provides tax credits and rebates to developers, meaning there is no direct government tax expenditure, and as a result, the policy reduces tax liability for developers and investors, shifting the market's focus toward producing affordable housing.[53] The Low-Income Housing Tax Credit or LIHTC is a prime example of tax credits being used to address the supply-side issues relating to affordable housing. LIHTC provides tax credits to developers, who can then sell them to investors to raise equity, which aligns with the economic logic behind tax incentives to lower the marginal cost of production and encourage affordable housing development by making the projects more financially attractive.[54] The policies hope to increase the lower-rent housing options by improving the supply of affordable housing units, especially to benefit essential service workers in mountain towns. However, potential drawbacks behind the tax incentive include the potential for inefficiencies by reducing tax revenue, and developers often only receive benefits after a project's completion, which can dilute the incentive's effectiveness.[55]

The Montana Housing Tax Credit Program builds on the Federal Low-Income Housing Tax Credit (LIHTC) program by supplying tax credits to incentivize the development and improvement of low-income housing units. As a tax policy intended to support affordable housing, the program offers two credit types: competitive 9% tax credits, distributed based on the priorities of Montana's Qualified Allocation Plan, and 4% tax credits distributed to maintain and improve existing affordable housing units.[56] The program has been successful in establishing affordable housing. In a 2022 report, Economist Derek Sheehan estimated that for every $1 lost in revenue, the credit program leveraged $2.69 in both public and private residential investment spending in the state economy. Further benefits mentioned include the expectation for the tax program to increase LIHTC units in Montana by 41%. Through further analysis, the Montana Housing Tax Credit Program is also expected to increase educational outcomes for children, like the chance of children receiving higher education in their lifetime. Another impact is the program's reduction of the number of cost-burdened households by 386 and statewide saving of low-income households eligible for LIHTC housing of $1.86 million a year.[57]

Colorado has also adopted the use of tax credits or rebates to address affordable housing availability. Colorado's Senate Bill 24-002 authorizes local governments to create property tax incentive programs. These programs can utilize tax credits or rebates to incentivize desirable affordable housing actions like converting short-term rentals to long-term rentals, which offers a more stable outcome for local workers.[58] More specifically, the act stipulates that incentive programs must address an "area of specific local concern", defined as a use of real property deemed necessary for preserving residents' welfare, such as access to housing. The economic impact of the act will depend on how local governments implement the bill and design their tax incentives program, but the flexibility allows local governments to address unique housing challenges and authorizes the use of necessary resources to address housing needs and potentially increase housing supply and community needs.[59]

Tax-funded programs that allocate government resources through grants, loans, and subsidies provide direct financial support to affordable housing projects, ameliorating housing-shortage problems in mountain towns. Unlike tax credits and rebates, which reduce a government's tax revenue by decreasing developers' and investors' tax liability, these programs involve direct and immediate government expenditure to support affordable housing projects in the hope of correcting market failures. This approach can stimulate housing production, but it requires ongoing government budgets, subjecting communities to potentially inefficient uses of tax resources if budgets are mismanaged.[60]

Colorado's House Bill 23-1304, also known as Proposition 123, is a leading example of a policy that leverages tax revenue to fund affordable housing support programs. The proposition permits Colorado to allocate 0.1% of its yearly state income tax towards housing efforts, equating to hundreds of millions of dollars in funding annually. Under the proposition, the State Affordable Housing Fund was established, and its funds are split between 40% towards the Affordable Housing Support Fund, administered by the Colorado Department of Local Affairs (DOLA), and 60% towards the Affordable Housing Financing Fund, overseen by the Office of Economic Development and International Trade. In the fiscal year 2023-2024, the fund collected $160 million and allocated about $27 million of it toward developing affordable homes in rural and resort communities, creating 685 new affordable housing units as a result.[61]

Colorado's Proposition 123 intends to address the affordable housing issue in a few ways. Given that the housing affordability issue is primarily a supply-side issue, the proposition requires local governments to increase their affordable housing stock by 3% per year for two years, reaching a total increase of 6% by December 31, 2026.[62] On the demand side, the proposition hopes to assist homeownership by allocating 50% of the Affordable Housing Support Fund's resources towards homeownership support, which involves down-payment assistance.[62] For the 2024-2025 fiscal year, $60 million in funding is available with a maximum of $70,000 per housing unit seeking development.[63] The proposition also seeks to support the homeless by allocating 45% of the Support Fund's resources toward services that help people experiencing homelessness.[62] Because geographical constraints are often a limiting factor for the housing supply in mountain towns, the Land Banking Program will aid local governments and nonprofit organizations in the maintenance and acquisition of lands for affordable housing. The program works by distributing grants to governments and forgivable loans to some nonprofits, creating a relatively low-stakes environment that encourages housing availability goals; forgivable loans are essentially grants, as either a part or the entire loan principal does not need to be repaid if certain requirements, like affordable housing goals, are met.[64]

These market-based policies are a step toward greater housing affordability, but they are not flawless. Policies that increase tax-based funding for affordable housing programs must

fall back on the core economic principle of allocating scarce resources towards areas that bring the highest marginal benefit per dollar, as greater efficiency typically drives more total benefit to communities. To do this, policies should prioritize low-income individuals, households, and essential service workers who will typically benefit more from an increase in financial support and are critical to the function of mountain towns.[65] There is a strong argument for some level of government intervention to address the housing shortage because land scarcity is an exaggerated problem. In mountain towns, continuously rising popularity suggests supply is unlikely to outpace or keep up with demand, making some government intervention arguably necessary to address the housing market failure and recognize community goals.[66]

Federal housing programs like the LIHTC and Housing Voucher Program provide affordable housing support, but are ultimately flawed. The LIHTC addresses affordable housing issues in mountain towns by providing a financial incentive for low-income rental housing, but ultimately falls short due to the program being unable to consistently construct and maintain affordable homes for low-income individuals. Also, the LIHTC is often overwhelmed as excessive demand will frequently outpace available resources, and as a result, the LIHTC falls short of additional state funding.[67] In this respect, the Montana Housing Tax Credit Program is a success as it reinforces the LIHTC, creating a more effective program. On the other hand, the Housing Voucher Program addresses the rising rental prices problem, which causes financial struggle for low-income households that rely on renting for housing. The Housing Voucher Program subsidizes low-income and at-risk individuals; however, this policy only treats the symptoms of the problem and does not address the supply side of the problem, which contributes to the high rental prices due to the housing shortage. Without sufficient supply-side policies that increase the affordable housing stock, demand-side policies like the Housing Voucher Program will become unsustainable as rental prices increase further. This issue is already beginning, as recently, tenant-based vouchers have received increased funding yet have failed to keep up with rising demand, leaving only one in four eligible homes able to receive support.[68,69]

Proposition 123 in Colorado, which authorizes 0.1% of state tax revenue to fund affordable housing projects, has been successful but not without its shortfalls. The program has a stable source of funding and is expected to generate $320 million by the end of the 2024-2025 year, and the program will likely continue to generate hundreds of millions of dollars a year, but the sustainability of its revenue depends on future tax revenue. The program has become present across Colorado's mountain towns as 200 of Colorado's 336 jurisdictions are eligible to apply for funding. A prominent example where funding has brought affordable housing is in the town of Frisco, where a $5 million grant purchased land beneath a multistory building, and the land is planned to be developed into workforce housing.[70] In Craig, Colorado, $2.7 million was granted to support 96 low and middle-income rental apartments.[71] However, Proposition 123 falls short in some areas. The program requires local governments to increase their affordable housing stock by 3% each year for the next three years.[62] Given geographic constraints, this rule can be more limiting to mountain towns than the program's funding is constructive, discouraging support. Another shortfall is the program's ability to prioritize true low-income individuals and households. The program authorizes funding for projects that support residents making between 60% and 100% of the median income in the project area. However, the median income is calculated including investment earnings, which means in small-population, affluent mountain towns like resort communities, the median income can be widely skewed, causing inefficient resource allocation and goals to be under-prioritized.[61]

Market-based policies, such as tax incentives and government-funded programs, present both benefits and drawbacks. Tax credits like the LIHTC and initiatives such as Montana's Housing Tax Credit Program offer a market-driven solution that encourages private sector involvement without direct government expenditure. These programs provide flexible solutions, but the notable drawback is the reliance on reduced tax revenue, which can strain public resources and create inefficiencies if tax incentives are not allocated properly. As a result, such tax policies require constant revision to ensure resources are being used efficiently. On the other hand, tax-funded programs like Colorado's Proposition 123 take a more direct intervention by using state funds to support affordable housing projects. This can bring quicker results, but once again is reliant on tax revenue and budget constraints. While market-based policy tools can be effective, their success depends on the government's balance of economic efficiency and social equity.

Command-and-Control Policies:

Command-and-control policies take a more direct approach to addressing housing affordability through government regulations and mandates that aim to control market outcomes. These policies often relate to rental prices and control, zoning laws, and development standards. These policies are designed to curb rising costs and regulate the tradeoff between the ease and quality standards of housing production. However, the effectiveness of some policies, like rent control, is often debated. Some economists argue that rent control and other regulations will lead to an overall reduction in the housing supply and instead argue that market-based approaches are more effective.[72,73]

Land use and developmental regulations are other types of policies that play a significant role in the supply of affordable housing. Land use and developmental regulations are command-and-control policies as they provide standards and rules that the market must comply with. Many land-use and developmental regulations are designed to ensure public safety and meet aesthetic and environmental goals; however, regulations also increase the marginal costs for developers by adding layers of complexity, time, and expenses.[74] Regulations can contribute to the explanation behind housing demand outpacing supply and driving up prices, especially in U.S. mountain towns where there is high popularity and geographical and environmental factors influence difficulty in land use. Policies that relax zoning and land-use restrictions can help encourage affordable

DOI: 10.36838/v7i10.84

housing development by easing the complexity of the production process. [75]

The Forest Service Flexible Housing Partnerships Act is another example of a zoning and land-use policy that seeks to support mountain towns across the United States by increasing the amount of available land for development. The act seeks to combat the scarcity of land in mountain towns that is contributing to a housing shortage by leasing some of the underutilized administrative sites owned by the Forest Service to organizations and programs that will utilize those lands for the benefit of mountain towns.[76] Specifically, section 8623 of the 2018 Farm Bill, titled "Authorization for Lease of Forest Service Sites," explains that by leasing out underutilized administrative lands to local governments or private entities, housing development can be fostered while simultaneously generating revenue for the Forest Service. However, to prevent the overuse of the Forest Service's lands, leased land is limited to 10 undeveloped areas that are no more than 40 acres each. [77]

The Forest Service Flexible Housing Partnerships Act, a part of the 2018 Farm Bill, creatively addresses the affordable housing challenge in Western U.S. mountain towns and seeks to benefit both the community through affordable housing and the Forest Service by generating revenues from leasing out underutilized lands. Given that lands can be leased for up to 100 years, the act has the potential to provide a longer-term solution to the land scarcity issue that restricts the supply of housing in mountain towns.[78] Many mountain towns are surrounded by federal lands and forests, and this act provides the possibility to access parts of those underutilized federal lands while simultaneously funding the Forest Service and supporting Forest Service workers.[79]

The act found success in Summit County, Colorado, where the White River National Forest agreed to lease 11 acres of its land to the Dillon Work Center Administrative Site for 50 years. The lease, approved on September 29, 2024, provided the opportunity for up to 117 affordable homes and other neighborhood infrastructure, providing housing security for local workers and lower-income residents.[80] The lease set a national precedent for partnerships between the Forest Service and affordable housing projects; however, future success is uncertain. The Forest Service Flexible Housing Partnerships Act is a part of the larger 2018 Farm Bill, which requires renewal every 5 years, and in 2023, its renewal date was extended to September 30, 2024.[81] Upon its renewal and after debate about government funding, the decision was made to continue funding the Farm Bill until December 20, 2024, but the decision left out some Farm Bill actions, like the authority of the Forest Service to lease some of its lands to local governments for affordable housing.[82] Although the Forest Service Flexible Housing Partnerships Act has an uncertain future, it's a unique opportunity to support affordable housing, and the success in Summit County makes it a policy that should be considered in the future.

Another command-and-control policy tool is short-term rental regulations, but they are controversial given the circumstances of many mountain towns. High tourism in mountain towns drives up housing and rental prices and shifts the market towards short-term rentals, resulting in existing community members being crowded out.[83] These conditions must be addressed to ensure a sustainable future for mountain towns, but the resulting challenge is managing the tradeoff between economic growth, driven dominantly by tourists, and equitable community goals, like the protection of people at risk of being priced out of their communities.[83] Short-term rental properties provide housing for tourists who interact with the economy by shopping at local stores and restaurants and engaging in tourist-related activities; they also generate tax revenue for governments. The concern is that if short-term rental properties increase without the total housing stock increasing, there will be less housing for the essential workers and long-time residents in the mountain-town communities.[84] While short-term rental properties have gained controversy for exaggerating the housing shortage in mountain towns, the problem is more complex, as short-term rental properties do not always make up a significant amount of a mountain town's housing stock. Even in areas where short-term rental properties do make up a large proportion of the total housing stock, many of these mountain towns have always maintained a large proportion of short-term rental properties due to vacation being fundamental to their economies.[84] Nonetheless, some mountain towns have adopted short-term rental regulations in an attempt to ameliorate the affordable housing shortage.

In Colorado's ski town, Steamboat Springs, short-term rental control has been established to support lower-income workers and people being displaced or forced to live in cramped conditions. Short-term rental control is simple; it sets a cap on the number of short-term rentals allowed in an area or prohibits the practice altogether. In Steamboat Springs, short-term rentals have been banned in the majority of the mountain town, and a 9% tax has been imposed on the practice to raise funds for affordable housing projects. The policy works by dividing Steamboat Springs into three zones.[85] Zone A, marked green on maps, where there is no restriction on short-term rentals. Zone B, marked yellow on maps, is divided into six subzones that determine the number of short-term rentals authorized. Zone C, marked in red on maps, prohibits short-term rentals. Of the three zones, zones A and B make up the smallest area, while Zone C makes up the vast majority of the mountain town.[86]

The tax, coupled with the major ban on short-term rental properties, has brought controversy to the community. On one hand, the policy supports lower-middle-income workers who are at risk of being priced out of the market and left to live outside the town with long commutes. On the other hand, tourists and some business owners are frustrated with the policy. Tourists rely on short-term rentals and contribute millions of dollars annually to mountain towns through spending, meaning the reduction in available housing for them will hurt local businesses.[87] The tax imposed on the short-term rental industry is another point of controversy, as some businesspeople are afraid Steamboat Springs' government will discourage tourists who support commerce and, as a result, damage the economy.[87] But, the short-term rental tax is expected to bring in millions of dollars annually to support affordable housing; in

2024, an estimated $14.8 million in revenue was generated by the short-term rental tax.[88]

In the context of the broader debate between market-based solutions, like the tax incentives in Montana, and more direct government intervention, command-and-control policies are sometimes preferred by the general public because they demonstrate a clear and straightforward effort toward resolving economic issues.[89] In contrast, market-based policies may appear too lenient, and are often criticized for increasing taxes on the consumer or using tax resources inappropriately. That said, command-and-control policies fall short in their flexibility. Traditional regulation is strict, typically treating different companies under the same assumptions, while market-based policies that often leverage financial incentives encourage firms to respond given their unique positions, resulting in more efficient economic outcomes.[89]

Command-and-control policies are a traditional form of market regulation. Unlike market-based policies, which rely on incentives to encourage and discourage specific actions, command-and-control policies establish rules that demand compliance. While market-based policies prioritize flexibility and cost efficiency, making them more economically favorable, command-and-control policies ensure specific decisions align with standards, but often at the cost of economic efficiency. Command-and-control policies not only include rules that prohibit actions but also include deregulation and the authorization of new actions. Although command-and-control policies may sacrifice efficiency, they ensure stability and protection in certain critical areas that require attention, making them highly useful depending on the circumstances.

Hybrid Policies:

Hybrid policies combine elements of market-based and command-and-control policies to influence desirable social outcomes. They provide the government with a powerful and flexible public-policy tool as they balance economic efficiency through market-based approaches by seeking cost efficiency while establishing the predictability of traditional regulatory standards.[90] These types of policies often come in the form of increased regulation or deregulation in an area while also creating a market incentive to further encourage specific actions and allow for greater economic efficiency without the use of strict rules. A common example of a hybrid policy is the cap-and-trade system for carbon emissions and environmental protection. The system establishes emission limits, representing a command-and-control policy, while establishing a pollution permit trading system, aligning with market-based ideals.[91] Critics of hybrid policy approaches often suggest they fail to reach maximum economic efficiency by limiting pure market forces through regulations. However, hybrid approaches retain the ability to address multiple policy objectives at once, encouraging desirable actions while mitigating the risks associated with pure market or regulatory approaches.

Deed restrictions are another example of a hybrid policy tool used to address affordable housing. Deed restrictions seek to allow homeowners to build some equity while preserving the affordability of home prices throughout the future.[92] Deed restrictions are legal constraints attached to a property's deed and embody command-and-control ideals by allowing local governments to directly regulate the housing market and preserve affordability. Deed restrictions limit the appreciation value of houses but attempt to create a separate, controlled market for affordable housing that is protected from the broader real estate market forces and ensure more affordable housing prices even during periods of high demand and limited supply.[92] From a market-based perspective, many deed restriction programs leverage financial incentives to encourage homeowners to adopt the restriction in exchange for financial benefits.[92]

Municipalities Vail and Mountain Village in Colorado have adopted deed restriction programs with unique features in an attempt to preserve affordable housing for local workers and lower-income households. In 2021, Mountain Village launched a deed restriction program called Your Equity Support (YES). The program offers homeowners 15-20% of their property's appraised value, up to $200,000, in exchange for a deed restriction being placed on the home. The deed restriction ensures that homes are being rented to local workers who work at least 1,560 hours annually. The YES program is unique because it does not set an appreciation, income, or household size cap, unlike traditional deed restriction programs that set appreciation caps.[93] Vail, Colorado, takes a similar approach by purchasing deed restrictions from willing homeowners, allowing local governments to ensure deed-restricted homes are owned by a local worker in Eagle County who works at least 30 hours a week.[94] Since its establishment in 2017, Vail has invested $12.5 million in purchasing 1,050 title deeds.[95]

Deed restrictions ensure affordable housing for some, but can be unsustainable and unfavorable for others' unique situations, creating a dynamic that challenges their effectiveness and universal application. Deed restrictions are often cited to create market distortions and can be difficult to implement widely as they hinder homeowners' flexibility, making them unfavorable for some. In mountain towns, where there is limited land for new housing construction, deed restrictions can be effective as they quickly convert existing market-rate homes to deed-restricted, affordable homes.[96] Depending on the circumstances, deed restrictions are also more cost-effective compared to building new affordable units, and they allow local workers and lower-income households to have greater availability of homes in the housing market.[97] While deed restrictions ensure long-term housing, they can be unpopular and hard to implement due to their restrictions on the flexibility of homeowners who have deed-restricted property. Public policy programs attempt to make up for this through financial incentives that support homeowners financing the property, and for many, the additional benefit from the funding outweighs the cost of losing some flexibility. The other fundamental concern with deed restrictions is that, despite their short-term effectiveness in providing affordable housing in areas with limited land for new housing, deed restrictions do not address an overall housing shortage, causing them to be ineffective in some circumstances.

Land-use and development regulation is another hybrid policy type. While more often a command-and-control policy, in special cases, the policy type can be coupled with financial

DOI: 10.36838/v7i10.84

incentives to reward or penalize a specific use or acquisition of land. Land use policies that rely purely on regulations demand adherence to standards, making it more difficult for projects to be approved, as they require development applications and approval. While these regulations can discourage projects, the addition of market-based incentives addresses this problem by encouraging people to pursue projects despite regulations, allowing for economic activity without compromising standards.[98] The nature of command-and-control policies also allows policymakers the opportunity to reduce regulations and provide financial incentives, which can be highly encouraging by making projects easier to initiate and more affordable.

Colorado's HB24-1152 is a hybrid policy that combines elements of both command-and-control and market-based approaches to improve housing affordability through accessory dwelling units. Accessory dwelling units, or ADUs, are small, independent housing units situated on the same land as a primary residence. On the command-and-control side, the policy mandates changes to zoning laws in cities and counties that meet certain criteria. In doing so, more ADUs will be allowed in areas where single-family homes are already permitted, increasing housing unit density and availability. On the market-based side, the policy provides financial grants to local governments; $5 million is allocated to offset ADU developmental costs, and another $8 million to facilitate affordable loans, interest rate buy-downs, and down-payment assistance.[99] The economic logic behind this policy is to increase the housing supply by making ADU construction easier and more financially feasible.

The municipality of Durango, Colorado, has adopted the hybrid approach to addressing housing affordability through ADUS. Durango has relaxed building restrictions to encourage more ADU development and increase the affordable housing unit stock. The relaxation of building requirements includes a reduction in parking requirements and eligible plot size, and the dimensional standards for the ADUs are less stringent, allowing for larger ADUs to be constructed. Durango believes more ADU construction will be instrumental in increasing affordable housing, as more than half of ADU owners who responded to surveys charge less than $1,000 a month, an affordable option for local workers.[100] As for a market-based policy, the city of Durango is providing $8,000 rebates for owners who rent their ADUs to local workers who work at least 32 hours a week in La Plata County and use the ADU as a primary residence. The $8000 comes from a City Council allocated funding of $80,000 that is administered on a first-come come-first first-served basis.[101]

Colorado's efforts to leverage ADUs to increase affordable housing provide the possibility to benefit both ADU owners and affordable housing recipients. In a joint 2022 survey by the City of Boulder's Housing and Human Services Department and the Planning and Development Services Department, it was revealed that the majority, 68%, of ADU owners view the primary benefit of ADUs as the supplemental income they provide.[102] From this survey, 64% of respondents claim their ADUs are used for long-term rentals, and on average these ADU rentals were priced at $1,626 in 2022, a 21% increase from 2017, a slower increase compared to the overall housing in Colorado which grew by 27% in cost from 2017 to 2022.[102] ADU owners benefit from the additional income, while the community benefits from additional affordable housing stock for individuals and small families. The survey data also supports the idea that ADUs provide long-term housing for many people, and ADU owners who prioritize supplemental income are incentivized to rent to long-term tenants, given a stable income. Long-term rentals are important as they prioritize housing security for essential workers and community members who are at greater risk compared to short-term visitors and tourists.

Despite their dual benefits, ADUs in Colorado are unlikely to entirely solve the affordable housing challenge. In the same 2022 survey by the City of Boulder, only 23% of respondents reported they would be willing to pursue the construction of an additional ADU if permitted.[102] This reveals that command-and-control policies, deregulation, and financial incentives have limited effects. That said, using ADUs to fight the affordable housing problem in mountain towns does not need to be a silver bullet to have positive impacts on communities.

The other shortfall with ADUs is their true affordability. While ADUs provide a more affordable option compared to traditional housing, their relative affordability can be deceiving as it masks the fact that their rent is still cost-burdening to many Coloradans. Rent that makes up more than 30% of one's income is considered cost-burdening.[103] In a 2021 study with the University of Minnesota, it was revealed that 47% of Colorado households made less than $75,000 in income, meaning 47% of Colorado households could afford a maximum monthly rent of $1,875, assuming they spend no more than 30% of their $75,000 income.[104] This value presents a dangerous reality, as many Coloradans who make less than $75,000 are still financially burdened by the average ADU price. Using 2023 data, the Bell Policy Center estimates that among Colorado households earning below 30% of the Area Median Income (MLI), 33% make less than $75,000, placing them in a financially insecure position where ADUs are not affordable.[105] ADUs do not provide financial stability for many low-income Coloradans who require affordable housing policy the most, so from some perspectives, Colorado's efforts to increase ADUs are unsuccessful in providing truly affordable housing for the lowest-income Coloradans. However, simultaneously, increasing the number of ADUs does provide an increased stock of housing units that are more affordable, and while they are not perfectly affordable, they are still a better option than traditional housing, making the Colorado-ADU policies imperfect but a better alternative than doing nothing.

Hybrid policies are a versatile approach to addressing complex challenges that policymakers often face. They combine the structure and certainty of command-and-control regulation with the flexibility and efficiency of market-based incentives to balance economic efficiency with standards deemed necessary. Deed restrictions provide long-term affordability by preserving a part of the housing market for local workers and lower-income individuals. Though they limit flexibility, policymakers can incentivize their adoption with financial rewards, making

them favorable options for some. The notable shortfall of deed restrictions is their failure to increase the total housing supply, which makes them unsustainable solutions, given some circumstances. ADU policies address the total housing supply by increasing the housing stock through zoning reform and financial incentives to encourage construction. While their relative affordability remains a concern for the lowest-income individuals, they provide a more affordable option for renters and benefit homeowners by creating an additional flow of income. Hybrid policies may not entirely resolve the affordable housing crisis, but their ability to manipulate tradeoffs and create incentives makes them a valuable tool to bridge the gap between traditional regulation and market dynamics.

Comparing the Different Policies:

Each policy takes a different approach to targeting affordable housing and leverages a different amount of resources, leading to varying results. For brief comparisons, the state-planned Montana low-income housing tax credit policy is estimated to affect 122 housing units annually through encouraging tax credits.[57] For Colorado's Proposition 123, the funding done in the 2024 fiscal year is planned to support 8,049 housing units through lower-interest loans, grants, and other direct subsidies.[106] In the Flexible Forest Housing Act, its application in the Whiteriver National Forest is set to support up to 117 additional affordable homes in the area by opening up more available land for affordable construction.[80] Finally, in Steamboat Springs, 2,342 housing units are under short-term rental regulations as of August 2024, an attempt to manage price inflation for housing rentals.[107] The varying outcomes are due to multiple factors like funding, addressable community size, and intensity of encouragement Colorado's Proposition 123 is producing some of the largest impacts because of its high funding and extensive reach, policies like the Forest Service Flexible Housing Partnerships Act take more specific and slow-moving effort to encourage affordable housing.

■ Conclusion

Economic Sustainability and Constructive Steps Forward:

The future of mountain towns involves the challenge of balancing opportunities for growth while working within the constraints of geography and preserving local environments and culture. Tourism is foundational to the economy and its growth as it draws visitors seeking idyllic landscapes, natural wonders, and recreational activities. While lucrative, the seasonal nature of tourism can cause harm as periods of overwhelming demand can strain local infrastructure, degrade the environment, and foster economic inequality between affluent visitors and residents and workers. Tourism and population growth are essential to the growth of mountain towns, but they also affect housing affordability and job stability, and threaten the sustainability of the environmental quality that originally attracted visitors to the towns.

A recent and developing issue in mountain towns is the shortage of affordable housing. The trend is the result of difficulties in balancing demand with a constrained supply. Geographic constraints and environmental limitations make new construction challenging and expensive despite increasing demand, raising the financial incentive. Affluent residents and second-home owners put upward pressure on housing and rental prices, leaving prices out of an affordable range for some essential workers and longtime community members. Even with demand-side policies like housing vouchers and programs aimed at supporting low-income renters, these efforts are unsustainable and fall short without addressing the supply side of the challenge by increasing the affordable housing stock. Traditional market forces have proven to be slow and insufficient in correcting the disparity between supply and demand due to constraints, making the housing challenge in mountain towns particularly unique and challenging.

Efforts to address these challenges have led to innovative policy approaches that combine both market-based incentives and command-and-control regulations. Programs like the Low-Income Housing Tax Credit (LIHTC) and state-specific initiatives such as Colorado's Proposition 123, which allocates tax revenue to support affordable housing projects, offer potential pathways to support affordable housing throughout the future. Although market-based policies provide financial incentives, maintain flexibility, and more closely align with classical economic thinking, they often rely greatly on public funding and require careful oversight to ensure efficient resource usage. Command-and-control policies take a more direct approach through rules like zoning and construction standards. Such policies rely on regulation to ensure desired outcomes or deregulation of existing policies to encourage actions that foster social and economic goals.

Achieving sustainable economic development means allowing mountain towns to prosper now and in the future. To achieve this goal, policymakers should consider thoughtful policies that encourage development, attraction, and increased revenues for local businesses without neglecting the health of local ecosystems, local workers, and community goals. While mountain towns present a unique set of constraints that make sustainable goals formidable, they also provide distinct and memorable experiences and lifestyles for families and individuals and offer natural wonders that cannot be found elsewhere.

■ Acknowledgments

Thank you to my parents for supporting my passion and engaging in discussions on this topic. I also thank my mentor, Fernando Loayza-Jordan from Indigo Research, for his support and inspiring self-confidence within me. I attest that the ideas and writing in this paper are entirely my own.

■ References

1. Kirellos, G. *10 Must-Visit Historic Mountain Towns in the US.* The Collector. https://www.thecollector.com/must-visit-historic-mountain-towns-us/ (accessed 2025-02-05).
2. Brajcich, K. *Vail, Colorado: What a Sustainable Mountain Resort Destination Looks Like*. Sustainable Travel International. https://sustainabletravel.org/vail-what-sustainable-mountain-destination-looks-like/ (accessed 2025-01-29).

DOI: 10.36838/v7i10.84

3. ORR. *New Data Shows Outdoor Recreation is a $1.2 Trillion Economic Engine, Supporting 5 Million American Jobs.* Outdoor Recreation Roundtable. https://recreationroundtable.org/news/new-data-shows-outdoor-recreation-is-a-1-2-trillion-economic-engine-supporting-5-million-american-jobs/ (accessed 2025-01-29).
4. Dent, M. Powder and Profits: The Economics of Ski Resorts. *The Hustle*. January 20, 2024. https://thehustle.co/powder-and-profits-the-economics-of-ski-resorts (accessed 2025-01-29).
5. Lawson, M. *Housing in recreation-dependent counties is less affordable*. Headwaters Economics. https://headwaterseconomics.org/equity/housing-affordability-recreation-counties/ (accessed 2025-02-05).
6. Averill, G. The 24 Best Mountain Towns in the U.S., Ranked. *Outside Online*. January 6, 2022. https://www.outsideonline.com/adventure-travel/destinations/north-america/best-mountain-towns-us/ (accessed 2025-02-05).
7. Körner, C.; Urbach, D.; Paulsen, J. Mountain Definitions and Their Consequences. *Alp. Bot.* **2021**, *131* (2), 213–217. https://doi.org/10.1007/s00035-021-00265-8.
8. Tiwari, P.; Bhagwati, J. Mountains: Definition, Classification, Significance, Marginality, Vulnerability, Environmental Changes and Sustainable Development. https://sunraise.sfu-kras.ru/sites/default/files/Learning%20materials.pdf.
9. McGranahan, D. *Natural Amenities*. Economic Research Service. https://www.ers.usda.gov/topics/rural-economy-population/natural-amenities/ (accessed 2024-08-15).
10. Sullivan, W.; Rees, M. *The Mountain Migration Report 2021*; Northwest Colorado Council of Governments, 2021. https://www.nwccog.org/wp-content/uploads/2021/06/Mtn-Migration-Report-FINAL.pdf.
11. NWCCOG/CAST Community Assessment Project. https://www.nwccog.org/wp-content/uploads/2024/05/Presentation-Deck-Insights-Collective-5.2.24-final.pdf.
12. McKee, S. Colorado Reported to Have Highest Share of "Remote Workers" in US. *The Gazette*. May 23, 2024. https://gazette.com/news/colorado-reported-to-have-highest-share-of-remote-workers-in-us/article_e6968608-a5b4-55e9-9bc7-947cfcd9179f.html.
13. American Trails. *Trails Make Economic Sense*. American Trails. https://www.americantrails.org/resources/trails-make-economic-sense (accessed 2024-08-22).
14. United Nations Environmental Programme (UNEP). *As plastic pollution piles up, mountain tourism reaches a crossroads*. UN Environment Programme. https://www.unep.org/news-and-stories/story/plastic-pollution-piles-mountain-tourism-reaches-crossroads (accessed 2024-08-15).
15. Song, H.; Dwyer, L.; Li, G.; Cao, Z. Tourism Economics Research: A Review and Assessment. *Ann. Tour. Res.* **2012**, *39* (3), 1653–1682. https://doi.org/10.1016/j.annals.2012.05.023.
16. United Nations Tourism. *Mountain Tourism*. UN Tourism. https://www.unwto.org/mountain-tourism.
17. McMahon, E.; Selzer, L. *About Gateway Communities*. Center for Sustainable Destinations. https://web.archive.org/web/20120421054103/http://travel.nationalgeographic.com/travel/sustainable/about_gateway_communities.html (accessed 2025-06-10).
18. Andonovski, V. Tourism and Mountain Development, 2014. https://www.fao.org/fileadmin/templates/mountain_partnership/images/Vlatko_ANDONOVSKI.pdf.
19. Grau, K.; Schultz, M.; Brown, G.; Weddell, M. *Understanding Visitor Recreation Experiences on Montana's Public Lands*; Montana Business Quarterly, 2023. https://www.bber.umt.edu/pubs/Seminars/2023/EconRpt2023.pdf.
20. Dean Runyan Associates. *The Economic Impact of Travel in the Estes Park Local Marketing District*; Simpleview, 2022. https://assets.simpleviewinc.com/simpleview/image/upload/v1/clients/estespark/EstesPark_Final_2021_Sept_6_2022_0be64323-e686-4f66-a0d2-798b32b7ff44.pdf.
21. Dean Runyan Associates. *The Economic Impact of Travel in Wyoming*; Travel Wyoming, 2023. https://industry.travelwyoming.com/research/annual-reports/?utm_medium=email&utm_source=sharpspring&sslid=MzM3NbI0MjIxMzC0BAA&sseid=MzIzMTUzsTC2MAcA&jobid=f7bb581b-b442-45aa-8644-226ad5ce0ea2 (accessed 2024-08-14).
22. Leaver, J. The State of Utah's Travel and Tourism Industry, 2024. *Kem C Gardner Policy Inst*. 2024.
23. National Ski Areas Association (NSAA). *State of the U.S. Ski Industry*; National Ski Areas Association, 2024. https://nsaa.org/webdocs/Media_Public/IndustryStats/Econ_impact_infographic_2024.pdf.
24. Saunders, S.; Easley, T. *Climate Change in the Headwaters Water and Snow Impacts*; The Rocky Mountain Climate Organization, 2018.
25. RRC Associates. *North Carolina Ski Areas Association 2022/2023 Economic Value Analysis*; North Carolina Ski Areas Association, 2023. https://www.goskinc.com/wp-content/uploads/2023/12/2022-23-NCSAA-Economic-Value-Analysis-Final-Report-10.10.23.pdf.
26. Iversen, E. K.; Grimsrud, K.; Lindhjem, H.; Navrud, S. Mountains of Trouble: Accounting for Environmental Costs of Land Use Change from Tourism Development. *Tour. Manag. Vol. 102*.
27. Zeng, L.; Yi Man Li, R.; Nuttapong, J.; Sun, J.; Mao, Y. Economic Development and Mountain Tourism Research from 2010 to 2020: Bibliometric Analysis and Science Mapping Approach. *MDPI Sustain*.
28. Kyara, V. C.; Rahman, M. M.; Khanam, R. Investigating the Environmental Externalities of Tourism Development: Evidence from Tanzania. *Heliyon* **2022**, *8* (6), e09617. https://doi.org/10.1016/j.heliyon.2022.e09617.
29. Gillingham, K.; Sweeney, J. Market Failure and the Structure of Externalities. **2010**.
30. Tian, X.; Gao, W.; Liu, Y.; Xu, M. Secondary Resource Curse's Formation and Transmission Mechanism Based on Environmental Externality Theory. *Resour. Conserv. Recycl.* **2020**, *161*, 104958. https://doi.org/10.1016/j.resconrec.2020.104958.
31. United Nations Tourism. *Tourism Seasonality*; World Tourism Organization (UNWTO), 1997. https://doi.org/10.18111/9789284401505.
32. Koldus, K. M. Affordable Housing in Mountain Resort Towns: Policy Recommendations for June Lake, Mono County, CA. *Mono County*. **2004**.
33. Solís, A. T.; Gil, G. L. The Impact of Tourism on the Labor Market of Local Service Industries in Barcelona. *Law Econ*. **2024**, *3* (7), 21–29.
34. Rawlings, *A. Mountain Town Migration: Understanding the Impacts of the Covid-19 Pandemic on Middle Neighborhoods in the Mountain West*; Joint Center for Housing Studies, 2023. https://www.jchs.harvard.edu/sites/default/files/research/files/harvard_jchs_mountain_town_migration_rawlings_2023.pdf.

35. Colorado Association of Realtors. *Local Market Update for January 2019, Pitkin County*; Colorado Association of Realtors, 2019. https://car-co.stats.showingtime.com/docs/lmu/2019-01/x/PitkinCounty?src=page.
36. Colorado Association of Realtors. *Local Market Update for January 2024, Pitkin County*; Colorado Association of Realtors, 2024. https://car-co.stats.showingtime.com/docs/lmu/2024-01/x/PitkinCounty?src=page.
37. Colorado Association of Realtors. *Local Market Update for January 2019, Grand County*; Colorado Association of Realtors, 2019. https://car-co.stats.showingtime.com/docs/lmu/2019-01/x/GrandCounty?src=page.
38. Colorado Association of Realtors. *Local Market Update for January 2024, Grand County*; Colorado Association of Realtors, 2024. https://car-co.stats.showingtime.com/docs/lmu/2024-01/x/GrandCounty?src=page.
39. Hedges, D. A. Who Gets to Live in the Mountains. Master's Thesis, Cornell University, 2023. https://ecommons.cornell.edu/server/api/core/bitstreams/b6ff7575-f3ad-4b0e-b148-0c5de69e0cb8/content.
40. Cole, A. The Mountain Lions: These Nine Cities Boomed in the COVID Era. *Full Stack Economics*. December 22, 2022. https://www.fullstackeconomics.com/p/the-rise-of-the-mountain-lion-economies (accessed 2024-08-14).
41. Lawson, M.; Smith, K. *Amenity Trap: How High-Amenity Communities Can Avoid Being Loved to Death*; Headwater Economics, 2023. https://headwaterseconomics.org/wp-content/uploads/2023HE-Amenity-Report-R3b-LOWRES.pdf.
42. Warsing, R. *Solving Colorado's Housing and Climate Crisis*. RMI. https://rmi.org/solving-colorados-housing-and-climate-crisis/ (accessed 2024-08-26).
43. Kenney, A. Colorado Has a Record-Breaking Affordable Housing Plan. Will State Lawmakers Push for Growth and Density, Too? *CPR News*. February 8, 2022. https://www.cpr.org/2022/02/08/colorado-affordable-housing-plan-growth-density/ (accessed 2024-08-24).
44 National Low Income Housing Coalition. The GAP: The Affordable Housing Gap Analysis 2021. https://nlihc.org/sites/default/files/gap/Gap-Report_2021.pdf.
45. Newcomer, J.; Resnick, P. *Factors Impacting Housing Affordability*; Colorado Futures Center, 2020. https://coloradofuturescsu.org/wp-content/uploads/2020/09/Housing-Unaffordability-Report_Executive-Summary_Fact-Sheets_Final.pdf (accessed 2024-08-24).
46. Dickhausen, J.; Weiler, S.; Report, R. Short-Term Rental Provisions and Colorado Mountain Housing Prices.
47. Ward, B. *What the New Wave of In-Migration Means for the State*; Montana Business Quarterly, 2023. https://www.bber.umt.edu/pubs/Seminars/2023/EconRpt2023.pdf.
48. Dineen, J. K. Tahoe's New Gold Rush: Bay Area Residents Fleeing Coronavirus Push Up Home Prices. *San Francisco Chronicle*. August 25, 2020. https://www.sfchronicle.com/bayarea/article/Tahoe-s-new-Gold-Rush-Bay-Area-residents-15511767.php (accessed 2024-09-02).
49. Wetzstein, S. Toward Affordable Cities? Critically Exploring the Market-Based Housing Supply Policy Proposition. https://doi.org/10.1080/10511482.2021.1871932.
50. Peace, J.; Ye, J. *Market Mechanisms: Options for Climate Policy*; Center for Climate and Energy Solutions, 2020. https://www.c2es.org/wp-content/uploads/2020/04/market-mechanisms-options-climate-policy.pdf.
51. Beiser-McGrath, L. F.; Bernauer, T.; Prakash, A. Command and Control or Market-Based Instruments? Public Support for Policies to Address Vehicular Pollution in Beijing and New Delhi. *Environ. Polit.* **2023**, *32* (4), 586–618. https://doi.org/10.1080/09644016.2022.2113608.
52. Schell, B. *Policy Approaches for Improving Housing Affordability and Availability*. Network for Public Health Law. https://www.networkforphl.org/news-insights/policy-approaches-for-improving-housing-affordability-and-availability/ (accessed 2024-10-14).
53. Local Housing Solutions. *Federal programs for affordable housing.* Local Housing Solutions. https://localhousingsolutions.org/fund/federal-programs-for-affordable-housing/ (accessed 2024-09-30).
54. Desai, M. A.; Dharmapala, D.; Singhal, M. *Investable Tax Credits: The Case of the Low Income Housing Tax Credit*; Working Paper; Harvard Business School, 2008. http://www.ssrn.com/abstract=1150302 (accessed 2024-09-22).
55. Roberts, B. *The Role of Tax Incentives in Affordable Housing*. National Housing Conference. https://nhc.org/the-role-of-tax-incentives-in-affordable-housing/ (accessed 2024-09-30).
56. Maciver, A.; Varns, K. E.; Jorgenson, C.; Soller, W.; Cestnik, M. *Supporting Low-Income Housing: The Montana Board of Housing and Its Role in Awarding Housing Tax Credits*; Legislative Audit Division. https://dailymontanan.com/wp-content/uploads/2024/06/23P-04-Low-Income-Housing-6-5-2024-1.pdf.
57. Sheehan, D. *An Assessment of the Potential Impacts of a Montana Low-Income Housing Tax Credit*; Bureau of Business and Economic Research, University of Montana, 2022. https://www.bber.umt.edu/pubs/econ/LowIncomeHousingTaxCreditImpact2022.pdf.
58. Colorado House Democrats. *New Laws to Spur Affordable Workforce Housing, Address Rising Property Insurance Rates in Rural, Mountain Communities Go Into Effect*. Colorado House Democrats. https://www.cohousedems.com/news/new-laws-to-spur-affordable-workforce-housing,-address-rising-property-insurance-rates-in-rural,-mountain-communities-go-into-effect (accessed 2024-10-14).
59. Frizell, L.; McCluskie, J.; Roberts, D. *Local Government Property Tax Credits Rebates*; 2024. https://leg.colorado.gov/bills/sb24-002.
60. Bundrick, J. *Tax Incentives and Subsidies: Two Staples Of Economic Development – Arkansas Center for Research in Economics*. Arkansas Center for Research in Economics. https://uca.edu/acre/2016/08/19/tax-incentives-and-subsidies-two-staples-of-economic-development/ (accessed 2024-09-30).
61. Tann, R. Has Proposition 123 — Colorado's Voter-Approved Affordable Housing Fund — Made a Difference in High-Cost Mountain Towns? *SummitDaily*. August 20, 2024. https://www.summitdaily.com/news/colorado-mountain-town-affordable-housing-fund/ (accessed 2024-09-16).
62. Colorado Department of Local Affairs. *Proposition 123 Commitment Filing Now Open (2-Year 6%)*. Proposition 123 Implementation. https://engagedola.org/prop-123.
63. Colorado Department of Local Affairs. *Prop 123 Homeownership Funding for FY2025*. Proposition 123 Implementation. https://engagedola.org/prop-123.
64. Colorado Housing and Finance Authority. *Land Banking*. Colorado Affordable Housing Financing Fund. https://coloradoaffordablehousingfinancingfund.com/land-banking/ (accessed 2025-02-25).

DOI: 10.36838/v7i10.84

65. Housing Colorado. *2024 Policy Priorities*. Housing Colorado. https://www.housingcolorado.org/page/policypriorities (accessed 2024-10-21).
66. Milton Jr., R. M. Resort Towns Need to Get Serious About Affordable Housing. *Governing*. February 28, 2024. https://www.governing.com/urban/resort-towns-need-to-get-serious-about-affordable-housing (accessed 2024-10-21).
67. Wilde, A.; West, A.; Hasstedt, K.; Alderman, C.; Visani, G.; Webster, S.; Whitham, A. *Colorado's Affordable Housing Crisis: It's Time For Strategic Investments*; Colorado Coalition for the Homeless, 2023. https://www.coloradocoalition.org/sites/default/files/2023-04/COAffordableHousingCrisis_StrategicInvestmentReport_CCH_2023.pdf.
68. Colorado Department of Local Affairs. *Housing Voucher Programs*. Colorado Department of Local Affairs Division of Housing. https://doh.colorado.gov/housing-voucher-programs (accessed 2024-10-21).
69. *Federal Budget and Spending | National Low Income Housing Coalition*. https://nlihc.org/federal-budget-and-spending (accessed 2024-10-21).
70. Clauer, D. *$5 Million Grant Awarded For Frisco Affordable Housing Project*. Breckenridge, Colorado Real Estate - Real Estate of the Summit. https://realestateofthesummit.com/summit-county-news/5-million-grant-awarded-for-frisco-affordable-housing-project/ (accessed 2025-02-25).
71. Craig Press. *Developers land $2.7M state grant to support Craig workforce apartments*. Craig, Colorado Economic Development. https://discovercraig.com/index.php/news/article/developers-land-2.7m-state-grant-to-support-craig-workforce-apartments (accessed 2025-02-25).
72. Valdez, R. *How Rent Control Makes Housing Less Affordable.* FREOPP. https://freopp.org/whitepapers/rent-control-paper/ (accessed 2024-10-14).
73. Rosen, K.; Van Dyke, D.; Bank, D.; Park, J. *The Effect of Rent Control on New Housing Supply: A Bay Area Case Study*; 2020.
74. Calder, V. *Zoning, Land-Use Planning, and Housing Affordability*. Cato Institute. https://www.cato.org/policy-analysis/zoning-land-use-planning-housing-affordability (accessed 2024-09-22).
75. Harris, C. *The Exclusionary Effects of Inclusionary Zoning: Economic Theory and Empirical Research*; Manhattan Institute, 2021. https://manhattan.institute/article/the-exclusionary-effects-of-inclusionary-zoning-economic-theory-and-empirical-research/ (accessed 2024-09-30).
76. Office of Senator Michael Bennet. *Bennet, Daines, Neguse Introduce Bill to Address Housing Shortage in Rural and Mountain Communities*. Michael Bennet, U.S. Senator for Colorado. https://www.bennet.senate.gov/public/index.cfm/2023/5/bennet-daines-neguse-introduce-bill-to-address-housing-shortage-in-rural-and-mountain-communities (accessed 2024-09-08).
77. United States Congress. *Section 8623 of the 2018 Farm Bill. Authorization for Lease of Forest Service Sites*; 2018. https://www.fs.usda.gov/sites/default/files/2018-farm-bill-section-8623.pdf.
78. Rep. Neguse, J. [D-C.-2. *Text - H.R.3562 - 118th Congress (2023-2024): Forest Service Flexible Housing Partnerships Act of 2023*. Library of Congress. https://www.congress.gov/bill/118th-congress/house-bill/3562/text (accessed 2024-11-04).
79. Office of Representative Joe Neguse. *Rep. Neguse Passes Bill to Address Housing Shortage in Rural & Mountain Communities Through Committee with Unanimous Support | Congressman Joe Neguse*. Representative Joe Neguse. https://neguse.house.gov/media/press-releases/rep-neguse-passes-bill-address-housing-shortage-rural-mountain-communities (accessed 2024-11-04).
80. U.S. Forest Service. *White River National Forest and Summit County sign Dillon Work Center lease*. U.S. Forest Service. https://www.fs.usda.gov/detail/whiteriver/news-events/?cid=FSEPRD1142052 (accessed 2024-11-04).
81. Zaks, L. *Government Funding and Farm Bill's Future Top Congress's September To-Do List*. National Sustainable Agriculture Coalition. https://sustainableagriculture.net/blog/government-funding-and-farm-bills-future-top-congress-september-to-do-list/ (accessed 2024-11-04).
82. Tann, R. U.S. Forest Service Loses Ability to Lease Land for Affordable Housing after Congress Fails to Renew Measure in Government Funding Bill. *Steamboat Pilot & Today*. September 28, 2024. https://www.steamboatpilot.com/news/u-s-forest-service-loses-ability-to-lease-land-for-affordable-housing-after-congress-fails-to-renew-measure-in-government-funding-bill/ (accessed 2024-11-04).
83. Albuquerque, H.; Quintela, J. A.; Marques, J. The Impact of Short-Term Rental Accommodation in Urban Tourism: A Comparative Analysis of Tourists' and Residents' Perspectives. *Urban Sci*. **2024**, *8* (3), 83. https://doi.org/10.3390/urbansci8030083.
84. Dubetz, A.; Horton, M.; Kesteven, C. Staying Power: The Effects of Short-Term Rentals on California's Tourism Economy and Housing Affordability. *Milken Inst*. **2022**.
85. City of Steamboat Springs. *Municipal Tax | Steamboat Springs, CO - Official Website*. City of Steamboat Springs. https://steamboatsprings.net/119/Municipal-Tax (accessed 2025-02-05).
86. City of Steamboat Springs. *Rules & Regulations | Steamboat Springs, CO - Official Website*. City of Steamboat Springs. https://www.steamboatsprings.net/1361/Rules-Regulations (accessed 2025-02-05).
87. Bedayn, J. V*acation towns limit short-term rentals amid housing crisis*. PBS News. https://www.pbs.org/newshour/nation/vacation-towns-limit-short-term-rentals-amid-housing-crisis (accessed 2025-02-05).
88. Steamboat Pilot & Today staff report. Steamboat Allocates $5 Million in Short-Term Rental Tax Revenue to Housing Authority for Cottonwoods Project. *Steamboat Pilot & Today*. January 30, 2025. https://www.steamboatpilot.com/news/steamboat-allocates-5-million-in-short-term-rental-tax-revenue-to-housing-authority-for-cottonwoods-project/ (accessed 2025-02-05).
89. Zhang, B. Market-Based Solutions: An Appropriate Approach to Resolve Environmental Problems. *Chin. J. Popul. Resour. Environ*. **2013**, *11* (1). https://doi.org/10.1080/10042857.2013.777526.
90. EPA. *Economic Incentives*. United States Environmental Protection Agency. https://www.epa.gov/environmental-economics/economic-incentives (accessed 2024-11-13).
91. Brooks, S. A.; Keohane, N. O. The Political Economy of Hybrid Approaches to a U.S. Carbon Tax: A Perspective from the Policy World. *Rev. Environ. Econ. Policy* **2020**, *14* (1), 67–75. https://doi.org/10.1093/reep/rez022.
92. Chen, J. F. Deed Restrictions And Affordable Home Ownership. *Harv. John M Olin Fellows Discuss. Pap. Ser*. **2023**.

(93) Town of Mountain Village. *YES Deed Restriction Program*. Town of Mountain Village. https://townofmountainvillage.com/community/housing/yes-deed-restricted-program/ (accessed 2024-11-13).
(94) Office of Policy Development and Research. *Vail, Colorado: The Vail InDEED Program Provides Deed-Restricted Workforce Housing in a Resort Market | HUD USER*. Office of Policy Development and Research. https://www.huduser.gov/portal/casestudies/study-081121.html (accessed 2024-11-13).
(95) Carrier, N. *Addressing Housing Challenges in Popular Seasonal Destinations*. Local Housing Solutions. https://localhousingsolutions.org/addressing-the-housing-crunch-in-seasonal-destinations/ (accessed 2024-11-13).
96. Park Place Finance. *What Do Deed Restrictions Mean for Property Investments?*. Park Place Finance. https://parkplacefinance.com/deed-restrictions/ (accessed 2024-11-13).
97. Horvath, A. *More mountain towns taking drastic measures to address housing shortages*. Rocky Mountain PBS. https://www.rmpbs.org/blogs/news/more-mountain-towns-taking-drastic-measures-to-address-housing-shortage (accessed 2024-11-13).
98. Kayden, J. S. Market-Based Regulatory Approaches: A Comparative Discussion of Environmental and Land Use Techniques in the United States. *Boston Coll. Law* **2022**.
99. Exum, T.; Mullica, K.; Weinberg, R.; Amabile, J. *Accessory Dwelling Units*.
100. Burney, C. Durango Seeks to Grow Workforce Housing with Rebate for Accessory Dwelling Units. *Durango Herald*. September 6, 2022. https://www.durangoherald.com/articles/durango-seeks-to-grow-workforce-housing-with-rebate-for-accessory-dwelling-units/ (accessed 2024-09-09).
101. City of Durango. *ADU Program Information*. City of Durango. https://www.durangoco.gov/850/ADU-Program-Information (accessed 2024-09-09).
102. City of Boulder. *Accessory Dwelling Unit Update Evaluation*; City of Boulder, 2022. https://bouldercolorado.gov/media/9924/download?inline.
103. National Low Income Housing Coalition. *About the Gap Report*. National Low Income Housing Coalition. https://nlihc.org/gap/about (accessed 2024-11-04).
104. Resnick, P.; Newcomer, J. *Housing Affordability Update*; Colorado Futures Center, 2023. https://www.coloradofuturescsu.org/wp-content/uploads/2023/08/CFC_Colorado_Housing_Affordability_Update_Summer2023.pdf.
105. Kniech, R. *Can ADUs Deliver on the Promise of Affordability for Colorado?*; Bell Policy Center, 2024. https://www.bellpolicy.org/wp-content/uploads/FINAL-Can-ADUs-Deliver-Robin-Kniech-04-2024.pdf.
106. Lifari, P.; Anderson, C.; Gamm, E. *Colorado's Housing Crossroads: Prop 123, State Investments, Market Challenges, and The Road Ahead*. https://www.commonsenseinstituteus.org/colorado/research/housing-and-our-community/colorados-housing-crossroads-prop-123-state-investments-market-challenges-and-the-road-ahead (accessed 2025-06-11).
107. Ballantyne, T. *Council reviews changes for short-term rental code in Steamboat Springs*. https://www.steamboatpilot.com/news/council-reviews-changes-for-short-term-rental-code-in-steamboat-springs/ (accessed 2025-06-11).

Author

Matthew Hagedorn is a junior at Aragon High School in San Mateo, California. He loves to ski with his friends and family and has a deep interest in economics and public policy. Matt plans to continue skiing, reading, and learning more about effective economic policy in college.

DOI: 10.36838/v7i10.84

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Impact of Substance-Induced Psychosis on Adolescents

Preethi Anbarasan

The Royal High School, East Barnton Avenue, Edinburgh EH4 6JP, UK; Preethi.a2211@gmail.com

ABSTRACT: Cannabis exposure during adolescence, the critical period of neurodevelopment, causes profound and often permanent neurobiological changes with negative long-term implications. These include reduced educational attainment, depression, anhedonia, anxiety, impulsive behavior, and problems with memory and attention. This review explores the effect of substance-induced psychosis (SIP) on adolescents, in particular adolescent cannabis users, who have an 8-fold higher incidence of psychosis than non-users, a vulnerability not seen in adults. Cannabis' influence on the endocannabinoid system (ECS) is of key importance for adolescents as CB1 receptor expression peaks between ages 14-17, increasing the tetrahydrocannabinol (THC) induced disruption of the endocannabinoid system. This has a profound effect on synaptic pruning and global excitation-inhibition balance, directly impacting brain development, in particular, dopaminergic systems. SIP remains challenging to diagnose due to overlap with other psychotic disorders, and misdiagnosis is common, with 25% of people diagnosed with PPD in emergency departments having SIP. Both SIP and schizophrenia's symptoms include ideas of reference, persecutory delusions, and auditory hallucinations. Positive and negative psychotic symptoms caused by persistent cannabis use disorders increase in severity over time. Cannabis users exhibit poorer symptomatic outcomes than those who use other illicit substances. They have an earlier age of onset of psychosis and poorer social functioning than alcohol misusers. Cannabis induced psychotic disorders have a higher risk of conversion to schizophrenia than alcohol induced psychosis. Meta-analysis reveals that cannabis-induced psychosis (CIP) had the highest rate of progression to schizophrenia (34%) than all other substances, while alcohol had the lowest (9%).[1] Cognitive behavioral therapy (CBT), motivational interviewing (MI), contingency management (CM), rewards programs, and familial involvement are effective when treating adolescents with cannabis use disorders. The Adolescent Community Reinforcement Approach (A-CRA) is also an effective treatment for cannabis abstinence. Recreational Cannabis use (CU) in adolescents could be reduced by banning cannabis marketing, limiting retail expansion, and increasing its price and legal age of purchase. Adolescents can access support through rehabilitation centers offering cannabis detoxification and non-pharmacological treatments. Given the impact on neurodevelopment and high psychosis conversion rates, protecting adolescents from CU during this critical developmental window represents a public health priority.

KEYWORDS: Translation Medical Sciences, Disease Detection and Diagnosis, Psychiatry, Substance Use.

Introduction

292 million people worldwide used illicit drugs in 2022, which has risen by 20% since 2012. Cannabis remains the most widely used drug across all ages, making up 78% of global illicit drug use, with 228 million users in 2022.[1,2] While cannabis usage rates appear stable in American adolescents (25.8% in 12th graders in 2024,[3] tetrahydrocannabinol (THC) potency has increased from 4% in 1995 to 15-20% today,[4] increasing the exposure for adolescents. Additionally, recreational cannabis use is now legalized in multiple areas, including Canada, Uruguay, and 27 jurisdictions in the US.[2] Daily cannabis users are now more common in the USA than daily alcohol users, and while the median drinker uses alcohol on 4-5 days per month, the median cannabis consumer uses the drug on 15-16 days per month.[5]

This widespread use is of particular concern for adolescents, who face unique neurodevelopmental vulnerabilities during a critical period of brain maturation. Illicit substance use may have longitudinal effects, including psychosis, defined as the presence of delusions, which are fixed false beliefs, and[6] hallucinations, which are sensory perceptions that occur in the absence of an actual external stimulus.[7] Substance-induced psychosis (SIP) is when substances such as cannabis, stimulants, or alcohol contribute to the onset of these psychotic symptoms.[7]

Cannabis intoxication behaves in a dose-responsive manner, meaning greater consumption results in greater intoxication and the potential for greater long-term side effects.[8] As shown in Figure 1, earlier initiation of cannabis is linked to a greater likelihood of developing psychosis; the cumulative incidence of psychotic disorders was eight times higher among adolescents who used cannabis than those who had not. Comparatively, Figure 1 displays a smaller difference in cumulative psychotic disorder incidence between young adults who had used cannabis and those who did not.[9] This age-dependent vulnerability demonstrates the significance of focusing on the adolescent demographic, as CU during this time period is a far stronger predictor of psychotic disorders for them than for adults. Early diagnosis and intervention are necessary to prevent disruption of the endocannabinoid system and subsequent long-term cognitive symptoms and psychiatric disease.

Cannabis use, however, is not the sole cause of SIP, as substances like alcohol and psychostimulants can also cause this. Social impacts of SIP include reduced educational attainment and employment acquisition. This review article analyzes the

DOI: 10.36838/v7i10.99

current academic literature on SIP, the biological influences on the brain, and long-term social implications with a particular focus on CU. In addition, this article provides a critical analysis of the current difficulties in the field of diagnosing SIP, the impacts of different substances on SIP severity, and treatment and intervention.

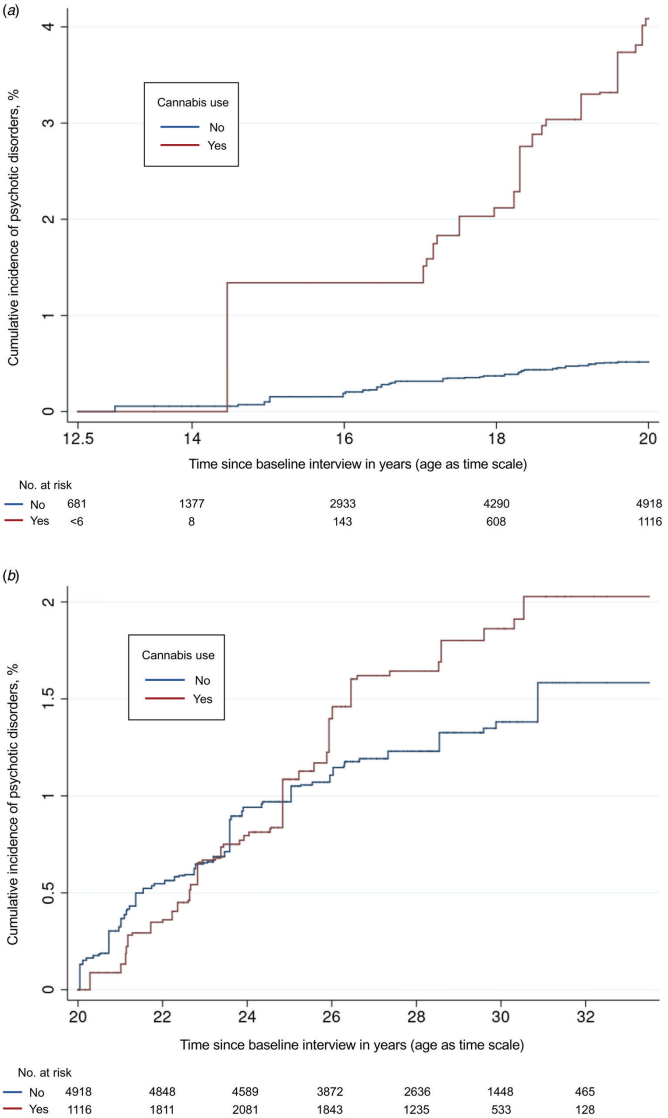


Figure 1: Comparative longitudinal study displaying that CU at adolescence is associated with a greater risk of developing psychotic disorders.[9] 15–20-year-olds using cannabis were associated with increased cumulative psychotic incidence than those not using it. This difference was less significant in 20–32-year-olds.

Biological Influences of THC on the Adolescent Brain:

To understand why adolescents show this increased vulnerability to psychosis, we must understand the normal function of the endocannabinoid system (ECS) and how THC disrupts this. The ECS plays a key role in memory, learning, reward, and pain pathways in the brain and consists of cannabinoid receptors and endocannabinoids (eCBs). There are two subtypes of cannabinoid receptors known as cannabinoid receptor type 1 (CB1R) and cannabinoid receptor type 2 (CB2R). CB1Rs are the most abundant G-protein-coupled receptors in the brain, with their highest density found in the hippocampus, basal ganglia, cerebellum, and neocortex.[10] Their expression peaks between ages 14-17, an age also associated with intensive synaptic pruning, a key neurodevelopmental process which removes weak synaptic connections. This "use it or lose it" process peaks between 16 and 20, especially in prefrontal brain regions associated with executive function and emotional regulation, which show up to a 40% reduction in synaptic density. THC exposure during this period leads to excessive pruning via overactivation of the ECS, potentially explaining the long-term cognitive effects of adolescent CU.[11]

The main role of the ECS is the regulation of neurotransmission via retrograde transmission at the synapse; this function is impaired by THC exposure. Figure 2 shows how, in normal ECS functioning, eCBs regulate the release of GABA and glutamate neurotransmitters, fine-tuning the activity of their synapses and consequently the balance of excitation and inhibition within the brain. Figure 3 shows how THC acts as a partial agonist of CB1R, subsequently leading to chronic downregulation of CB1R and significant disruption to the ECS.[12] This disruption reduces sensitivity to stress and reward[13], and also the release of GABA and glutamate. Glutamate is an excitatory neurotransmitter, meaning it makes neurons more likely to fire. GABA is an inhibitory neurotransmitter, meaning it causes neurons to be less likely to fire. Overstimulation of the ECS by THC creates an imbalance of excitatory-inhibitory neurotransmission in the brain by reducing inhibition more than excitation.[12]

Rodent studies have shown that this excitatory-inhibitory imbalance increases psychosis-like behaviours.[12] As shown in Table 1, CU in adolescence has numerous neurological effects. Reduced GABA release due to THC disinhibits dopaminergic neurons, as the neurons that would normally suppress their activity become less active. This reduces the amount of cortisol released in response to tension, which heightens feelings of bliss.[13] The disruption of these dopamine pathways (Figure 4) increases susceptibility to addiction and psychosis.[12] Early exposure to THC in adolescence results in a reduced release of dopamine in adulthood in response to stress and psychostimulants[14], showing the lifelong impact on brain function, CU in adolescence can cause this. Crucially, while the ECS normally functions at a synapse level over a timeframe of minutes, THC activates the ECS across the whole brain for hours at a time, preventing it from performing its normal role.

Table 1: Summary of the impact of adolescent THC exposure. Various neurotransmitter systems are affected during THC exposure, which increases the risk of psychosis.

Neurotransmitter	Consequences of adolescent THC exposure
Glutamate (excitatory)	Reduced executive function and increased risk of psychosis.
GABA (inhibitory)	Anxiety
Dopamine	Decreased cortisol, heightened feelings of bliss, increased susceptibility to addiction and psychosis
Endocannabinoids	Reduced sensitivity to stress and reward

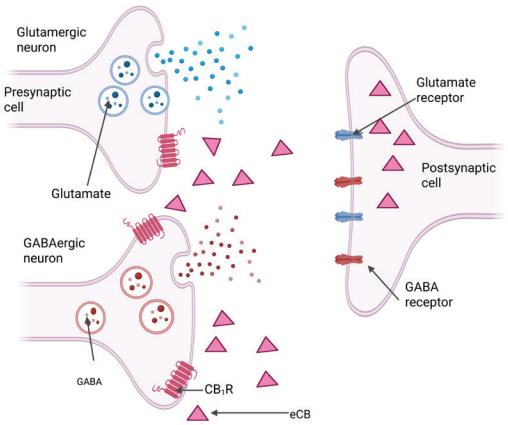


Figure 2: Normal Synaptic Transmission.[12] In the absence of cannabis, normal functioning of the ECS occurs. eCBs regulate the release of GABA and glutamate neurotransmitters, which create an excitatory-inhibitory balance in neurotransmission.

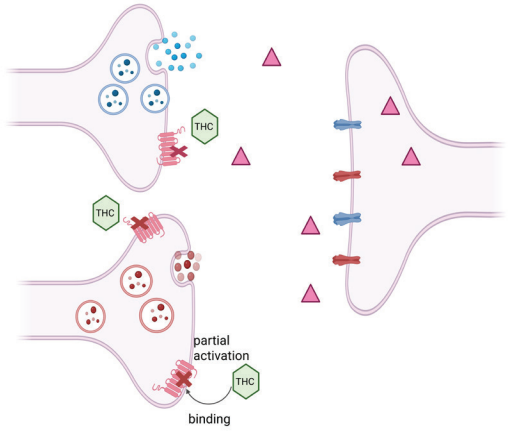


Figure 3: Effects of adolescent CU.[12] THC acts as a partial agonist of CB1R, downregulates CB1R, and reduces the release of GABA and glutamate. THC overstimulates the ECS, causing an excitatory-inhibitory imbalance.

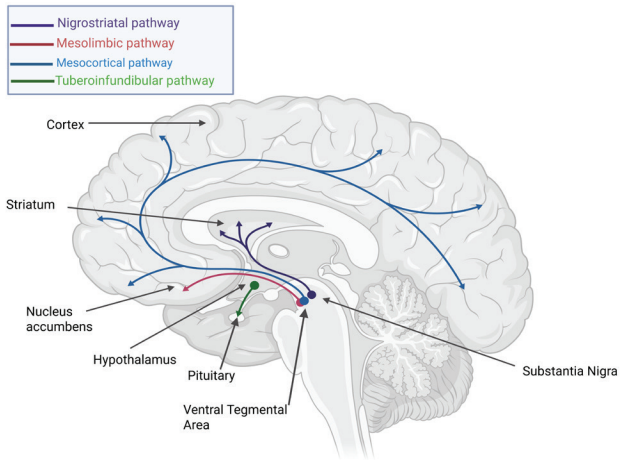


Figure 4: Dopamine pathways in the brain. THC exposure during adolescence disrupts dopamine pathways in the brain, which increases the risk of addiction and psychosis.

Long-term social implications:

Adolescent cannabis users are likely to suffer from poor cognitive function, showing problems with memory, attention, educational attainment, and,[15,16] impulsive behaviours.[15] A meta-analysis showed that CU in adolescence increased the relative risk of depression (37%), suicidal ideation (50%). In addition, the risk of suicide attempts was 3.46 times higher.[13] Regular CU in adolescence can also cause anhedonia and anxiety.[16] Persistent CU can result in increased criminal and risk-taking behaviors among adolescents.[17] Adolescents aged 14-15 display more evident increases than adults, proving that CU has a greater effect on younger demographics. However, in countries where cannabis remains illegal, this increased likelihood of criminal behavior may be due to regular users becoming connected to the illegal drug market and drug traders through their attempts to get access to cannabis. The user may then be influenced by people in the market who encourage participation in violent crimes.[17]

Studies show that CU lowers IQ and the ability for sustained focus.[18] This inhibition affects school performance and educational attainment. Adolescents who use cannabis tend to also be high-school dropouts and have lower attainment rates in their careers and education.[19] This highlights the need to reduce CU in adolescents through effective early diagnosis and intervention. The earlier the intervention, the lower the risk of experiencing significant adverse financial, health, and education outcomes.

Difficulty Diagnosing:

Given the significant neurobiological influence of CU in adolescence and its long-term social implications, accurate and early diagnosis of SIP is crucial. However, there are some limitations in diagnostic criteria and the overlap between SIP and schizophrenia, which can make diagnosis challenging. There are only minor changes between the current Diagnostic and Statistical Manual of Mental Disorders (DSM)-V criteria for SIP and the previous DSM-IV criteria for SIP. The DSM-V criteria now specify that individuals must have one or both of the following symptoms: delusions and/or hallucinations. It also now stipulates that the disturbance present in the individual must cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.[7] Due to the DSM-V criteria being more recent, the literature exploring its limitations is sparse. For this article, the DSM-IV criteria will be evaluated instead, which are relevant to the DSM-V as well.

The DSM-IV requires four criteria for an SIP diagnosis, as shown in Table 2. Diagnosing SIP can be uncertain, partly due to the current discrepancies in the DSM-IV criteria.

DOI: 10.36838/v7i10.99

Table 2: DSM-IV criteria for SIP.[20] The DSM-IV criteria require clinicians to assess whether hallucinations or delusions were present near or during a period of cannabis intoxication before diagnosing SIP. An SIP diagnosis also requires symptoms to persist in the absence of delirium, and these symptoms should not be better explained by a non-substance-induced psychotic disorder.

Criterion	Description
A	Prominent hallucinations or delusions (patients must lack the insight that substances cause their hallucinations).
B	i) symptoms from criterion A occurred during or within a month of intoxication or withdrawal ii) The disturbance is not caused by using medication
C	A non-substance-induced psychotic disorder cannot better explain the disturbance. Evidence of a non-substance-induced psychosis may include symptoms which manifest before the onset of substance use (or medication use), remain for a significant amount of time after acute withdrawal or severe intoxication has stopped or having symptoms over what would be predicted given the type, amount or duration of substance use. In addition, evidence such as a history of multiple non-substance-related episodes could rule out SIP.
D	The disturbance occurs not only during delirium.

Problems with diagnostic criteria:

In criterion A, the term 'prominent' is ambiguous, as there is no description of what severity or duration of symptoms would be considered 'prominent.' Criterion A's lack of information about other psychotic symptoms, including negative symptoms and disorganized speech, is an issue, as negative symptoms are common in SIP patients.[20]

Criterion A only lists hallucinations and delusions narrows the range of symptoms present in SIP and thus fails to account for the frequent discovery of many other symptoms in SIP. Criterion B's failure to focus on the criterion of withdrawal and intoxication also causes confusion. Psychotic episodes linked to SUD are not included in the SIP criteria if there is no intoxication. This disregards tolerance that is built by persistent substance use, which means individuals have a lower likelihood of being intoxicated. All these issues contribute to the underreporting of SIP, highlighting the need for improved criteria for increased accuracy in diagnosing.[20]

In emergency departments, when there is uncertainty, psychotic disorders tend to be reported as primary psychotic disorder (PPD) rather than SIP,[21] highlighting that a large proportion of patients fail to be diagnosed with SIP at the start of hospitalization as a result of the imprecision of DSM IV. As SIP is underreported, clinicians fail to identify and treat many adolescents with SIP, which may contribute to continuing CU into adulthood due to their increased dependence. Furthermore, failure to receive treatment correlates with poorer clinical and educational outcomes as cannabis use in adolescents interferes with the ECS during peak brain maturation.

There are currently different definitions for substance-induced psychosis across DSM-V and the International Classification of Diseases (ICD). A key difference between DSM 5 and ICD 10/11 is that DSM-V requires only the presence of psychotic symptoms after substance use for diagnosis, whereas ICD 10 and ICD 11 believe the sole presence of psychotic symptoms is not enough for a diagnosis. In ICD-10 and ICD-11, the symptoms should be significantly more severe than what is expected in the intoxication or withdrawal of the particular substance and dosage used.[22] These inconsistencies across criteria may cause underdiagnosis or misdiagnosis of SIP, highlighting the need for a standardized definition.

Overlap with schizophrenia:

When diagnosing SIP, we must consider not only the diagnostic requirements but also the similarities between SIP and schizophrenia. Schizophrenia is defined as "a heritable, complex, multi-dimensional syndrome with varying degrees of psychotic, negative, cognitive, mood, and motor manifestations,"[23] specifically two or more of the following symptoms: hallucinations, disorganized speech, delusions, and grossly disorganized or catatonic behavior for at least 6 months. At least one of the symptoms present must be delusions, hallucinations, or disorganized speech.[7] Symptoms that overlap between schizophrenia and substance-induced psychosis are ideas of reference, persecutory delusions, and auditory hallucinations.[24] Cannabis-induced psychosis (CIP) and schizophrenia possess overlapping genetic risk factors,[25] so accurate diagnosis cannot be made solely by looking at the individual's genes. Heritability estimates reveal 11% of the variance in lifetime CU is explained by measured genetic variants.[26] This overlap is problematic because the similarities between schizophrenia and CIP often make a differentiated diagnosis difficult. Individuals with schizophrenia may use illicit substances as well, thus further blurring the lines between schizophrenia and CIP.

Studies show that in emergency departments, 25% of people diagnosed with PPD had substance-induced psychosis.[21] The dangers of misdiagnosing SIP as PPD include the added burden of stigma, a low chance of clinicians correcting the diagnosis in the future,[21] with only 15% of misdiagnoses corrected within 6 months,[27] unmerited hospitalizations, and incorrect medication being prescribed. The misdiagnosis can lead to the use of antipsychotic medication and increase the risk of developing side effects like tardive dyskinesia, neuroleptic malignant syndrome, and diabetes.[21]

Treatments for adolescents with SIP, who are incorrectly diagnosed with schizophrenia, would focus on alleviating psychotic symptoms rather than specific treatments that address the root cause of their psychosis, that is, their substance use. Ineffective treatments stemming from an inaccurate diagnosis are particularly harmful for adolescents whose positive life outcomes depend on early detection and treatment. To maximise treatment impacts, it is important to identify the drug of choice. Common culprits of SIP are cannabis, alcohol, and psychostimulants, especially as these substances are known to impair critical neurological developments occurring during adolescence. The specific psychotic trajectories of CU, psychostimulants, and alcohol can also be compared to gain a better understanding of their impact and treatment.

Comparison with other illicit substances:

Persistent cannabis use disorder (CUD) is correlated with positive and negative symptoms, which worsen over time. CUD is the only substance use disorder in which symptomatic and functional decline occurs from year 1 to year 2 in a 2-year follow-up, compared with alcohol use disorder (AUD)

and psychostimulant use disorder. Even cannabis users who comply with medication have a greater likelihood of having a poor symptomatic outcome than other substance misusers. Continuing to use cannabis after the development of first-episode psychosis (FEP) can cause symptom levels to increase. In addition, there is a dose-dependent relation between psychotic symptoms and CU, meaning that the greater the amount consumed, the more severe the psychotic symptoms will be. The gradual decline in the condition of cannabis misusers may stem from intense, persistent CU and low compliance with medication at the beginning of the psychotic disorder.[8] This worsening trajectory highlights the significance of early interventions for cannabis users.

Psychostimulants can also induce SIP, and their addictive properties may make cessation difficult. In the prospective longitudinal cohort study by Plamondon, there was a higher rate of addiction in the 24th month among cocaine and amphetamine users when compared to cannabis and alcohol users. These findings indicate that stopping psychostimulant use may be more challenging than cessation of alcohol or cannabis. In this study, participation selection bias was reduced as it ensured that all consenting and qualified individuals within established catchment areas were able to take part. However, a limitation of this study is that it did not possess objective methods of assessing substance use, such as urine tests.[8]

Psychostimulant misusers with FEP tend to have higher rates of unemployment than FEP patients who use cannabis.[8] It is clear that psychostimulants have harmful effects on individuals and should receive intervention as early as possible. The impact of stimulant use on psychotic symptoms was less dose-dependent than cannabis, and it possessed differing trajectories for occasional and persistent use. Even though both cannabis and stimulant use affect dopaminergic transmission, which dysregulates and influences psychotic symptoms and psychotic relapse, CU appears to be involved to a lesser degree than stimulant use. Due to the direct effect on dopaminergic transmission, even the occasional use of stimulants is enough to aggravate the psychotic symptoms and psychotic relapse. The severity of psychotic symptoms achieved by occasional use of psychostimulants can only be matched by CU if it is frequent.[28] By focusing on alcohol use in comparison to CU, we can discern that different substances have trends and trajectories and better understand the need for stricter cannabis regulation.

Alcohol vs Cannabis:

Cannabis induces over-activation of the endocannabinoid system by cannabinoid receptor type 1 agonists such as THC. Chronic over-activation of endocannabinoids during adolescence changes brain maturation and can have a lasting impact on the adult brain.[29] Alcohol consumption increases the sensitivity of the mesocorticolimbic dopamine network, which consists of the mesolimbic and mesocortical dopamine pathways in the brain. This may lead to an increase in positive symptoms.[30]

CIP and alcohol-induced psychosis display differences in clinical characteristics in adolescents. These fundamental differences underlie distinct clinical trajectories. CU is correlated with more severe positive psychotic symptoms compared to alcohol, which has a greater association with anxiety.[31] CIP displays a higher conversion rate to schizophrenia than alcohol-induced psychosis.[22,32] A Scottish register-based cohort study of 3486 patients diagnosed with SIP in Scottish hospitals found CIP had a 21.4% risk of conversion to schizophrenia compared to alcohol-induced psychotic disorders with a 10.4% risk.[33] However, the data did not contain drug screening confirmation, and the data were collected from 1997 to 2012, meaning it may not be an accurate reflection of current substance use trends and treatment approaches.

In a Finnish register-based cohort study of 18,478 SIP cases, the risk of conversion to schizophrenia was significantly higher for CIP(46%) than for alcohol-induced psychosis(5%).[34] While both studies show a higher transition rate from CIP to schizophrenia, they are retrospective and therefore are more susceptible to bias than prospective studies. In the Finnish study, only 0.7% of 18,478 cases were diagnosed with CIP compared to 85.4% diagnosed with alcohol-induced psychosis. Cannabis is illegal in both Finland and Scotland. The large difference between alcohol-induced psychosis and CIP suggests the results are confounded by underreporting out of fear of facing legal repercussions. This selection bias from illegal cannabis status would artificially inflate relative risk ratios while underestimating absolute CIP prevalence. Prospective studies with biological verification remain urgently needed.

Those who misuse alcohol have a lower risk of showing symptoms found in schizophrenia, such as paranoia, hallucinations, and negative symptoms, than cannabis-induced psychotic disorder. Cannabis misusers have an earlier age of onset of psychosis and a shorter period of experiencing illness,[35] but their symptoms deteriorate over time[8]. In contrast, alcohol misusers usually have a later age of onset with higher anxiety levels and,[31] higher hospitalization rates.[35] Despite this, they have better social functioning than cannabis users.[31] These dissimilarities highlight the importance of substance-specific approaches in the treatment and interventions of SIP in adolescents.

Treatment, Early Detection, and Intervention:

The difficulty in diagnosing CIP, along with the detrimental effects of substance use at the peak of brain maturation in adolescents, creates a call to action to reduce CUD rates. Psychological interventions combined with social interventions have shown promise for adolescents with CUDs. Adolescent CU often manifests as a social behavior encouraged by peers, and therefore, interventions such as MI, CM with rewards for not using cannabis would be effective. Combining motivational enhancement therapy (MET) and CBT with a rewards program led to longer periods of cannabis abstinence in adolescents. Familial involvement in treatments also proved beneficial for adolescents. Interventions for young adults should involve building resilience to pain, protective circles of peers, and training for better coping skills.[36]

A-CRA is another effective method for reducing CU in adolescents by providing cannabis users with problem-solving,

DOI: 10.36838/v7i10.99

communication, anger management, and relapse prevention skills through 19 procedures. An added benefit of A-CRA is that it is more cost-effective than CBT and MET.[37] In addition, having brief check-up sessions improved abstinence rates, highlighting the importance of follow-up meetings over many years. Clinicians should design interventions so that they capture the interest of the age group they are targeting. For example, using computerized or mobile programs appeals to young adults to strengthen motivation and allow self-monitoring.[36] CUD interventions should be designed and selected for each demographic to address the different challenges each age group faces when attempting abstinence.

When considering pharmacological interventions, there are no FDA-approved medications for CUD, and off-label psychotropic medications have had minimal benefits, such as mitigating withdrawal symptoms. However, medications to ease prolonged withdrawal symptoms when utilized in combination with counselling and social support have proved to be beneficial.[36] Furthermore, early detections and interventions are extremely beneficial for adolescents as this improves clinical and functional outcomes.

Prevention – Canada and solutions:

There are solutions to the problem of increasing CUDs that do not involve banning cannabis entirely. Cannabis-related ED visits declined in Canada following the legalization of recreational cannabis with tight retail regulations. However, it increased at the time of cannabis commercialisation.[38] This suggests that to minimize the increase in cannabis use following the legalization, cannabis marketing should be banned, and the expansion of retail stores selling cannabis should be prevented.

Additionally, increasing the price of cannabis could deter people from purchasing cannabis regularly, which reduces its accessibility by making it unaffordable for many. Studies have proved this by the high tax on cigarettes, which serves to decrease cigarette purchases and use. Adolescents are two to three times more reactive to changes in the pricing of cigarettes than the overall population.[39] This suggests that raising the price of illicit substances has the most significant impact on adolescents, as they are likely to have low-paid or part-time jobs. In Canada, the legal age for purchasing cannabis is 18; therefore by raising the legal age to 21, will prevent many adolescents from being able to purchase cannabis. By increasing the age of initiation of cannabis, neurological developments will not be affected as much, which could improve educational outcomes for many adolescents.

Support for Adolescents with CUD:

Adolescents struggling with drug addiction can get help through rehabilitation centers. There is a range of different rehabilitation centers, each targeted at dealing with a specific substance. These rehabilitation centers could create a sheltered, supportive environment in which adolescents with CUDs undergo a cannabis "detox" while participating in counselling, yoga, and meditation. A recent study from Andra Pradesh has displayed an increase in cognitive function among young adults with CUDs after they participated in non-pharmacological rehabilitation such as meditation, group counselling, physical activities, and yoga during periods of cannabis abstinence.[40]

Online programs targeting CUDs are increasingly present in Austria, Belgium, the Czech Republic, Spain, the UK, Sweden, Germany, the Netherlands, and Estonia. Examples of these programs include VALIC, CANreduce, and SIBRT. In addition, Belgium, Romania, and the UK now have cannabis-specific clinics which offer clinical assessments, detoxification, short-term and long-term rehabilitation to improve treatment outcomes such as cannabis abstinence and increase the likelihood of psychosis remission. They also aim to lower levels of CU and increase contentment with life. However, there are few cannabis-specific clinics targeted at adolescents, and in-person clinics have limited coverage as they only offer face-to-face treatments in some cities.[41]

Barriers that prevent participation in treatments for CUDs include fear of stigma, believing these treatments are not needed, feeling unmotivated to attempt abstinence, and poor mental health. Other reported barriers include finding it hard to admit they need support, limited treatment availability, worries about confidentiality, cynicism about the treatment's effectiveness, and a desire to be self-sufficient.[42] To further support adolescents, we must encourage them to seek help for their CU. By displaying posters with contact information for rehabilitation programs in public places that adolescents frequent, such as schools, gyms, and shopping centers, we can ensure that they know help is available. Individual schools should ensure that students struggling with drug addiction have a non-judgmental, trusted adult they can look to for support, such as a guidance counsellor. Increasing the support available for substance use in primary care settings would allow for earlier detection and interventions.[43]

■ Discussion

This review has demonstrated the mechanistic pathways linking adolescent CU to psychosis, with 25-46%[1] of patients with SIP from cannabis showing conversion to schizophrenia, far over other substances. While previous work has focused on adult populations, this review explains some of the unique vulnerabilities in adolescents that heighten their risk in a manner qualitatively different from adults. The significant biological influence of adolescent CU on the brain through the binding of THC to CB1Rs in the ECS. This binding dysregulates neurotransmission and lowers stress and reward sensitivity. Because CB1R expression peaks during adolescence, THC has a greater impact on adolescents than on other age groups. Specifically, THC lowers eCB levels in the ECS and reduces the release of GABA and glutamate, creating an excitatory-inhibitory imbalance which increases the risk of developing psychosis. Dopamine pathways are also disrupted, increasing vulnerability to addiction. These effects have long-term ramifications, including poorer cognitive function, educational attainment, and increased risk of depression, which can negatively impact the life outcomes of adolescent cannabis users.

The DSM-IV criteria have limitations that make diagnosing CIP challenging. Its limitations include its dismissal of psychotic symptoms when the individual has insight and a lack of information about negative symptoms, withdrawal, and intoxication. SIP has overlapping symptoms with schizophrenia, such as ideas of reference, persecutory delusions, and auditory hallucinations. These facts make it difficult to differentiate between the disorders and provide accurate diagnoses.

By comparison with alcohol, this review identifies that CU has a more detrimental effect on clinical and functional outcomes, while alcohol use has a clearer link to anxiety. Psychotic symptoms associated with CUD get worse over time, and CUD had the greatest severity of positive symptoms, while AUD had the least. Cannabis users tended to have an earlier age of onset of psychosis than alcohol users, although they experienced shorter periods of illness and fewer hospitalisations. CIP was more likely to convert to schizophrenia than alcohol induced psychosis. Despite these findings from Finnish and Scottish cohort studies being significant, the studies are retrospective, so the data is more vulnerable to bias than if they were prospective. The studies found many fewer CIP than alcohol-induced psychosis diagnoses, which may be due to underreporting out of fear of legal repercussions, as cannabis is illegal in both countries. Different substances have different trends and trajectories. Cannabis users have a poorer symptomatic outcome than other substance users, although psychostimulant use has lower rates of abstinence than cannabis. However, the study that identified this finding is limited by its lack of objective methods to measure substance use. Cannabis users' exacerbation of psychotic symptoms over time can be attributed to persistent, severe CU and non-compliance with medication at the start of psychosis.

Different demographics benefit from different treatments and interventions to promote abstinence from cannabis use. Off-label psychotropic medications do not have significant benefits, and there is no FDA-approved medication for CUD. A combination of MET, CBT, and a rewards program, as well as familial involvement in treatment, proved to be most effective for adolescents. Computerized programs that allow self-monitoring also appealed to this group. Abstinence rates improved when follow-up check-up sessions were incorporated into treatment. The legalisation of cannabis in Canada with tight regulations was not associated with large increases in CU, yet cannabis commercialisation was. Increasing the legal age of cannabis purchase from 18 to 21 in Canada would reduce the number of adolescents who use cannabis and increase the age of initiation for many.

This review has numerous limitations that should be acknowledged. Firstly, the rapidly evolving landscape of cannabis product development and its use and regulation creates a moving target. Studies performed on adolescent CU even 5-10 years ago may have limited relevance to an adolescent population now using cannabis concentrates of 60-90% THC. Geographical variation in legality and regulation also creates different exposure contexts that may not be comparable. Studies linking substance use to psychosis have mostly focused on adult populations who have differing neurobiology from adolescents, meaning findings may not be directly applicable. The heterogeneity of cannabis products encompassing diverse cannabinoid compounds, consumption methods, and usage habits complicates the interpretation of studies that typically specify cannabis use only. Finally, the absence (for ethical reasons) of any causal studies or trials of CU makes untangling premorbid vulnerability from the effects of cannabis difficult. In addition, cannabis users often use other psychoactive drugs, which analysts struggle to control for. Future research utilizing techniques such as Mendelian randomization may overcome some of these causal limitations, though it should also be noted that the Bradford Hill criteria for causation are largely met.

Conclusion

Adolescent CU has significant neurobiological effects, causing long-term social implications and increased risk of psychosis. By focusing on early detection and intervention, clinical and functional outcomes can be significantly improved. CB1R expression peaks in adolescence, making CU during this period particularly harmful, partly due to excessive synaptic pruning as well as negative cognitive effects that persist into adulthood.

Many critical research questions are outstanding, ranging from causality to biomarker development to the lack of specific treatment offered to affected adolescents. The unique vulnerability of adolescents to the increased potency of THC and the trend towards reduced regulation of their sale to adolescents represents a potential health crisis. Immediate action through information campaigns, adolescent access restriction, and investment in treatment infrastructure and research should be prioritized.

Future research should aim to perform longitudinal studies on the treatment outcomes of adolescents with CIP to identify interventions to maintain long-term cannabis abstinence. In addition, there should be more cannabis-specific rehabilitation clinics tailored to the adolescent population. Countries offering face-to-face support should increase their national coverage to maximise their impact on adolescents by making treatment easily accessible. Policymakers should aim to regulate CU by limiting retail expansion of recreational cannabis, banning cannabis marketing, raising the legal age of purchase, and increasing its price in countries where it is legal.

Acknowledgments

I would like to thank my research mentor, Barbara Biney, for her guidance throughout my research.

References

1. Murrie, B.; Lappin, J.; Large, M.; Sara, G. Transition of Substance-Induced, Brief, and Atypical Psychoses to Schizophrenia: A Systematic Review and Meta-Analysis. *Schizophr. Bull.* **2020**, *46* (3), 505–516. https://doi.org/10.1093/schbul/sbz102.
2. *UNODC World Drug Report 2024: Harms of world drug problem continue to mount amid expansions in drug use and markets*. United Nations : Information Service Vienna. https://unis.unvienna.org/unis/en/pressrels/2024/uniscp1177.html (accessed 2025-01-25).

DOI: 10.36838/v7i10.99

3. *Reported use of most drugs among adolescents remained low in 2024.* National Institutes of Health (NIH). https://www.nih.gov/news-events/news-releases/reported-use-most-drugs-among-adolescents-remained-low-2024 (accessed 2025-06-02).
4. Abuse, N. I. on D. *Cannabis Potency Data | National Institute on Drug Abuse (NIDA)*. https://nida.nih.gov/research/research-data-measures-resources/cannabis-potency-data (accessed 2025-06-02).
5. Caulkins, J. P. Changes in Self-Reported Cannabis Use in the United States from 1979 to 2022. *Addiction* **2024**, *119* (9), 1648–1652. https://doi.org/10.1111/add.16519.
6. Kapur, S. Psychosis as a State of Aberrant Salience: A Framework Linking Biology, Phenomenology, and Pharmacology in Schizophrenia. *Am. J. Psychiatry* **2003**, *160* (1), 13–23. https://doi.org/10.1176/appi.ajp.160.1.13.
7. Glasheen, C.; Batts, K.; Karg, R.; Hunter, D. Impact of the DSM-IV to DSM-5 Changes on the National Survey on Drug Use and Health.
8. Ouellet-Plamondon, C.; Abdel-Baki, A.; Salvat, É.; Potvin, S. Specific Impact of Stimulant, Alcohol and Cannabis Use Disorders on First-Episode Psychosis: 2-Year Functional and Symptomatic Outcomes. *Psychol. Med.* **2017**, *47* (14), 2461–2471. https://doi.org/10.1017/S0033291717000976.
9. McDonald, A. J.; Kurdyak, P.; Rehm, J.; Roerecke, M.; Bondy, S. J. Age-Dependent Association of Cannabis Use with Risk of Psychotic Disorder. *Psychol. Med*. **2024**, *54* (11), 2926–2936. https://doi.org/10.1017/S0033291724000990.
10. Zhang, M.; Chen, T.; Lu, X.; Lan, X.; Chen, Z.; Lu, S. G Protein-Coupled Receptors (GPCRs): Advances in Structures, Mechanisms and Drug Discovery. *Signal Transduct. Target. Ther*. **2024**, *9* (1), 88. https://doi.org/10.1038/s41392-024-01803-6.
11. Germann, M.; Brederoo, S. G.; Sommer, I. E. C. Abnormal Synaptic Pruning during Adolescence Underlying the Development of Psychotic Disorders. *Curr. Opin. Psychiatry* **2021**, *34* (3), 222–227. https://doi.org/10.1097/YCO.0000000000000696.
12. Bara, A.; Ferland, J.-M. N.; Rompala, G.; Szutorisz, H.; Hurd, Y. L. Cannabis and Synaptic Reprogramming of the Developing Brain. *Nat. Rev. Neurosci.* **2021**, *22* (7), 423–438. https://doi.org/10.1038/s41583-021-00465-5.
13. Fischer, A. S.; Tapert, S. F.; Lee Louie, D.; Schatzberg, A. F.; Singh, M. K. Cannabis and the Developing Adolescent Brain. *Curr. Treat. Options Psychiatry* **2020**, *7* (2), 144–161. https://doi.org/10.1007/s40501-020-00202-2.
14. Bloomfield, M. A. P.; Ashok, A. H.; Volkow, N. D.; Howes, O. D. The Effects of Δ9-Tetrahydrocannabinol on the Dopamine System. *Nature* **2016**, *539* (7629), 369–377. https://doi.org/10.1038/nature20153.
15. Lorenzetti, V.; Alonso-Lana, S.; Youssef, G.; Verdejo-Garcia, A.; Suo, C.; Cousijn, J.; Takagi, M.; Yücel, M.; Solowij, N. Adolescent Cannabis Use: What Is the Evidence for Functional Brain Alteration? *Curr. Pharm. Des.* **2017**, *22* (42), 6353–6365. https://doi.org/10.2174/1381612822666160805155922.
16. Bloomfield, M. A. P.; Hindocha, C.; Green, S. F.; Wall, M. B.; Lees, R.; Petrilli, K.; Costello, H.; Ogunbiyi, M. O.; Bossong, M. G.; Freeman, T. P. The Neuropsychopharmacology of Cannabis: A Review of Human Imaging Studies. *Pharmacol. Ther.* **2019**, *195*, 132–161. https://doi.org/10.1016/j.pharmthera.2018.10.006.
17. Fergusson, D. M.; Horwood, L. J.; Swain-Campbell, N. Cannabis Use and Psychosocial Adjustment in Adolescence and Young Adulthood. *Addiction* **2002**, *97* (9), 1123–1135. https://doi.org/10.1046/j.1360-0443.2002.00103.x.
18. Trezza, V.; Cuomo, V.; Vanderschuren, L. J. M. J. Cannabis and the Developing Brain: Insights from Behavior. *Eur. J. Pharmacol.* **2008**, *585* (2–3), 441–452. https://doi.org/10.1016/j.ejphar.2008.01.058.
19. Volkow, N. D.; Baler, R. D.; Compton, W. M.; Weiss, S. R. B. Adverse Health Effects of Marijuana Use. *N. Engl. J. Med.* **2014**, *370* (23), 2219–2227. https://doi.org/10.1056/NEJMra1402309.
20. Mathias, S.; Lubman, D. I.; Hides, L. Substance-Induced Psychosis: A Diagnostic Conundrum. *J. Clin. Psychiatry* **2008**, *69* (3), 358–367. https://doi.org/10.4088/JCP.v69n0304.
21. Schanzer, B. M.; First, M. B.; Dominguez, B.; Hasin, D. S.; Caton, C. L. M. Diagnosing Psychotic Disorders in the Emergency Department in the Context of Substance Use. *Psychiatr. Serv.* **2006**, *57* (10), 1468–1473. https://doi.org/10.1176/ps.2006.57.10.1468.
22. Tandon, R.; Shariff, S. M. Substance-Induced Psychotic Disorders and Schizophrenia: Pathophysiological Insights and Clinical Implications. *Am. J. Psychiatry* **2019**, *176* (9), 683–684. https://doi.org/10.1176/appi.ajp.2019.19070734.
23. Tandon, R.; Nasrallah, H.; Akbarian, S.; Carpenter, W. T.; DeLisi, L. E.; Gaebel, W.; Green, M. F.; Gur, R. E.; Heckers, S.; Kane, J. M.; Malaspina, D.; Meyer-Lindenberg, A.; Murray, R.; Owen, M.; Smoller, J. W.; Yassin, W.; Keshavan, M. The Schizophrenia Syndrome, circa 2024: What We Know and How That Informs Its Nature. *Schizophr. Res.* **2024**, *264*, 1–28. https://doi.org/10.1016/j.schres.2023.11.015.
24. Dawe, S.; Geppert, L.; Occhipinti, S.; Kingswell, W. A Comparison of the Symptoms and Short-Term Clinical Course in Inpatients with Substance-Induced Psychosis and Primary Psychosis. *J. Subst. Abuse Treat*. **2011**, *40* (1), 95–101. https://doi.org/10.1016/j.jsat.2010.08.002.
25. Keramatian, K.; Levit, A. Caring for Youth with Co-Occurring Substance Use and Severe Psychiatric Disorders: Diagnostic Challenges and Clinical Implications. *J. Can. Acad. Child Adolesc. Psychiatry* **2023**, *32* (3), 202–208.
26. Pasman, J. A.; Verweij, K. J. H.; Gerring, Z.; Stringer, S.; Sanchez-Roige, S.; Treur, J. L.; Abdellaoui, A.; Nivard, M. G.; Baselmans, B. M. L.; Ong, J.-S.; Ip, H. F.; van der Zee, M. D.; Bartels, M.; Day, F. R.; Fontanillas, P.; Elson, S. L.; de Wit, H.; Davis, L. K.; MacKillop, J.; Derringer, J. L.; Branje, S. J. T.; Hartman, C. A.; Heath, A. C.; van Lier, P. A. C.; Madden, P. A. F.; Mägi, R.; Meeus, W.; Montgomery, G. W.; Oldehinkel, A. J.; Pausova, Z.; Ramos-Quiroga, J. A.; Paus, T.; Ribases, M.; Kaprio, J.; Boks, M. P. M.; Bell, J. T.; Spector, T. D.; Gelernter, J.; Boomsma, D. I.; Martin, N. G.; MacGregor, S.; Perry, J. R. B.; Palmer, A. A.; Posthuma, D.; Munafò, M. R.; Gillespie, N. A.; Derks, E. M.; Vink, J. M. GWAS of Lifetime Cannabis Use Reveals New Risk Loci, Genetic Overlap with Psychiatric Traits, and a Causal Influence of Schizophrenia. *Nat. Neurosci.* **2018**, *21* (9), 1161–1170. https://doi.org/10.1038/s41593-018-0206-1.
27. Caton, C. L. M.; Hasin, D. S.; Shrout, P. E.; Drake, R. E.; Domínguez, B.; First, M. B.; Samet, S.; Schanzer, B. Stability of Early-Phase Primary Psychotic Disorders with Concurrent Substance Use and Substance-Induced Psychosis. *Br. J. Psychiatry* **2007**, *190* (2), 105–111. https://doi.org/10.1192/bjp.bp.105.015784.
28. Kreis, I.; Lagerberg, T. V.; Wold, K. F.; Åsbø, G.; Simonsen, C.; Flaaten, C. B.; Engen, M. J.; Lyngstad, S. H.; Widing, L. H.; Ueland, T.; Melle, I. Behind the Heterogeneity in the Long-Term Course of First-Episode Psychosis: Different Psychotic Symptom Trajectories Are Associated with Different Patterns of Cannabis and Stimulant Use. *Schizophr. Res*. **2024**, *271*, 91–99. https://doi.org/10.1016/j.schres.2024.07.006.
29. Krebs, M.-O.; Kebir, O.; Jay, T. M. Exposure to Cannabinoids Can Lead to Persistent Cognitive and Psychiatric Disorders. *Eur. J. Pain* **2019**, *23* (7), 1225–1233. https://doi.org/10.1002/ejp.1377.

30. Kirli, U.; Binbay, T.; Alptekin, K.; Kayahan, B.; Elbi, H. The Relationship between Alcohol-Cannabis Use and Stressful Events with the Development of Incident Clinical Psychosis in a Community-Based Prospective Cohort. *Turk. J. Psychiatry* **2021**. https://doi.org/10.5080/u26410.
31. Cookey, J.; McGavin, J.; Crocker, C. E.; Matheson, K.; Stewart, S. H.; Tibbo, P. G. A Retrospective Study of the Clinical Characteristics Associated with Alcohol and Cannabis Use in Early Phase Psychosis. *Can. J. Psychiatry Rev. Can. Psychiatr.* **2020**, *65* (6), 426–435. https://doi.org/10.1177/0706743720905201.
32. Bramness, J. G.; Hjorthøj, C.; Niemelä, S.; Taipale, H.; Rognli, E. B. Discussing the Concept of Substance-Induced Psychosis (SIP). *Psychol. Med.* **2024**, *54* (11), 2852–2856. https://doi.org/10.1017/S0033291724001442.
33. Alderson, H. L.; Semple, D. M.; Blayney, C.; Queirazza, F.; Chekuri, V.; Lawrie, S. M. Risk of Transition to Schizophrenia Following First Admission with Substance-Induced Psychotic Disorder: A Population-Based Longitudinal Cohort Study. *Psychol. Med.* **2017**, *47* (14), 2548–2555. https://doi.org/10.1017/S0033291717001118.
34. Niemi-Pynttäri, J. A.; Sund, R.; Putkonen, H.; Vorma, H.; Wahlbeck, K.; Pirkola, S. P. Substance-Induced Psychoses Converting into Schizophrenia: A Register-Based Study of 18,478 Finnish Inpatient Cases. *J. Clin. Psychiatry* **2013**, *74* (1), e94-99. https://doi.org/10.4088/JCP.12m07822.
35. Di Paolo, M.; Calabrese, A.; Nosari, G.; Ciappolino, V.; Cirella, L.; Caldiroli, A.; Capuzzi, E.; Clerici, M.; Buoli, M. Clinical Variables and Peripheral Biomarkers Associated with Substance-Induced Psychotic Disorder: Differences Related to Alcohol, Cannabis, and Psychostimulant Abuse. *J. Pers. Med*. **2024**, *14* (3), 325. https://doi.org/10.3390/jpm14030325.
36. Ghafouri, M.; Correa da Costa, S.; Zare Dehnavi, A.; Gold, M. S.; Rummans, T. A. Treatments for Cannabis Use Disorder across the Lifespan: A Systematic Review. *Brain Sci.* **2024**, *14* (3), 227. https://doi.org/10.3390/brainsci14030227.
37. Mehr, N. K.; Lavasani, F. F.; Noroozi, A.; Farahani, H.; Gharraee, B. Effectiveness of the Adolescent-Community Reinforcement Approach for Treating Cannabis Use Disorder in Iranian Adolescents: A Randomized Controlled Trial. *Acta Psychol. (Amst.)* **2024**, *251*, 104604. https://doi.org/10.1016/j.actpsy.2024.104604.
38. Myran, D. T.; Pugliese, M.; Tanuseputro, P.; Cantor, N.; Rhodes, E.; Taljaard, M. The Association between Recreational Cannabis Legalization, Commercialization and Cannabis-Attributable Emergency Department Visits in Ontario, Canada: An Interrupted Time–Series Analysis. *Addiction* **2022**, *117* (7), 1952–1960. https://doi.org/10.1111/add.15834.
39. Bader, P.; Boisclair, D.; Ferrence, R. Effects of Tobacco Taxation and Pricing on Smoking Behavior in High Risk Populations: A Knowledge Synthesis. *Int. J. Environ. Res. Public. Health* **2011**, *8* (11), 4118–4139. https://doi.org/10.3390/ijerph8114118.
40. Goud, S. S.; Radhika, M. S.; Indla, V.; Kolli, N. Cannabis Use and Effect of Cannabis Abstinence on Cognitive Functioning in Young People — an Observational Case-Control Follow-up Study from Rehabilitation Centre in Andhra Pradesh. *Middle East Curr. Psychiatry* **2022**, *29* (1), 61. https://doi.org/10.1186/s43045-022-00228-7.
41. Hoch, E.; Murawski, M.; Ferri, M.; Feingold, D. Cannabis Use Disorder: An Overview of Treatment Approaches in Europe. *Eur. Arch. Psychiatry Clin. Neurosci.* **2025**, *275* (2), 315–326. https://doi.org/10.1007/s00406-025-01964-7.
42. Sherman, B. J.; McRae-Clark, A. L. Treatment of Cannabis Use Disorder: Current Science and Future Outlook. *Pharmacother. J. Hum. Pharmacol. Drug Ther.* **2016**, *36* (5), 511–535. https://doi.org/10.1002/phar.1747.
43. Ghafouri, M.; Correa da Costa, S.; Zare Dehnavi, A.; Gold, M. S.; Rummans, T. A. Treatments for Cannabis Use Disorder across the Lifespan: A Systematic Review. *Brain Sci*. **2024**, *14* (3), 227. https://doi.org/10.3390/brainsci14030227.

Author

Preethi Anbarasan is currently a fifth-year student at the Royal High School and is interested in the field of medicine

DOI: 10.36838/v7i10.99

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Genetic Networks in Cognition, Addiction, and Homeostasis

Riya Shukla

West Chester East High School, 450 Ellis Ln, West Chester, PA 19380, USA; 1riyashukla011@gmail.com
Mentor: Dr. Jorge Avila

ABSTRACT: Genetic networks regulate cognition, addiction, and homeostasis by modulating human brain activity. In this paper, we will analyze four genetic networks - the mesolimbic system, acetylcholine network, CREB-dependent gene regulatory network, and NF-κB networks. We will analyze their roles in cognition, addiction, and homeostasis in humans, and look into how these networks relate. We will then go deeper into NF-κB pathways and analyze each of its 5 subunits to derive the common areas across the human body that are influenced by the genes interacting with those subunits. These common areas would be in terms of KEGG (Kyoto Encyclopedia of Genes and Genomes) and REACTOME pathways, disease pathways, and other biological processes. These will be termed as just *biological processes*. This derivation will be laid out in the form of Python code that can be reused to derive those common areas using the latest data from the trusted data sources. I have derived the specific genes that interact with each of the 5 NF-κB subunits and have found the biological processes they enrich. Recognizing genetic networks and their associated biological processes may help identify areas for further research or discover potential therapeutics for cognitive, addictive, or homeostatic diseases.

KEYWORDS: Cellular and Molecular Biology, Genetics, Cognition, Addiction, Homeostasis.

Introduction

The human body has around 19,000 to 25,000 genes, but the total number of genetic networks is still unknown. Genetic networks are defined by genes that control a similar process and can be categorized by cell type or purpose. They influence specific properties of the body, including cognition, addiction, and homeostasis. Cognition is the brain's ability to process and react to information and includes attention, decision-making, memory, and speech. Genes code for proteins that promote cell communication and synaptic plasticity, allowing neurons to have stronger bonds and stronger neural pathways.[1-4] These neurons and neural pathways are often activated when one is learning. Thus, stronger connections between them promote stronger memory and quicker responses to new stimuli, creating stronger cognition. Addiction is considered a chronic brain disease that happens when one abuses the use of a substance and is commonly classified into three main stages: the binge/intoxication phase, the withdrawal/negative stage, and the preoccupation/anticipation phase.[5,6] It is identified as a deficit in reward and dopamine pathways.[2,7,8] Addiction leaves problems in reward-based learning and memory that are usually permanent. Homeostasis is defined as how your body can self-regulate through external changes or stimuli.[2] The process uses feedback systems to maintain the body and ensure that activities such as temperature regulation, pH balance, and the proper management of specific elements are properly regulated.

It is vital to understand how genetic networks function under specific circumstances and in certain contexts to maintain the health of the brain and the rest of the body. Sometimes, there are issues in the brain that cause diseases across the body. If we research genetic systems and their effects, we can find target areas in the biological processes that are commonly susceptible to the underlying genes. This paper overviews four important genetic networks- mesolimbic, acetylcholine, CREB-dependent, and NF-κB - and their effects on cognition, addiction, and homeostasis. It then contrasts their roles and demonstrates a way of finding relationships within the networks.

Methods

The genetic networks considered in this study are the Mesolimbic System, the Acetylcholine Network, the CREB-dependent gene Regulatory Network, and the NF-κB Pathways. Existing research around these four genetic networks has been studied to derive their influence areas. With the listed areas, one genetic network, NF-κB, is chosen to understand the dynamics of the impacted pathways, considering the genes involved. In my research, I have used publicly available relevant data and have created the necessary Python code to surface the top biological processes enhanced by the genes relevant to NF-κB pathways in humans. This is achieved via 2 broad steps. First, I derive the gene names that interact with the NF-κB subunits. We will get 5 sets of genes, as there are 5 subunits of the NF-κB family. I then find the biological processes where those genes play a statistically significant role. Each set of genes influences a set of biological processes, and we consider only those where the p-value is less than 0.05 to ensure a higher degree of statistical significance. Uniprot is being used here as the data source since it is an open-source library of data relevant to NF-κB subunits (listed as proteins) and the genes that interact with their dependencies. This library collates and maintains experimental data from several trusted sources like IntAct, MINT, STRING, etc., and allows an API-based interface, which means we can acquire the latest data from them every time we run the code. GProfiler is used to derive the top

DOI: 10.36838/v7i10.108

relevant biological processes enriched by the genes interacting with each of the NF-κB subunits above. KEGG and REACTOME are major sources of biological processes referred to in this case. The Python code for the above work has been made publicly available here: **https://github.com/1Riya-Shukla/NF-kB-Pathways**

Results

Mesolimbic System:

The mesolimbic system is an important dopaminergic genetic network. It plays a key role in the brain's dopaminergic reward pathway.[7,8] It is found in the midbrain and works to connect these brain regions to the ventral tegmental area (VTA). The VTA contains neurons that produce and send dopamine, a neurotransmitter, across the brain in reaction to external stimuli associated with pleasure.[2] This pathway drives individuals toward goals and behavior repetition, and may even strengthen connections between cues and rewards.[7,8] A study found that defeated rats showed increased phasic dopamine signaling, a short-term boost of dopamine release during aggressive interactions, suggesting a link between social stress and dopamine activity.[8] Thus, the mesolimbic system can perform cognitive tasks by using dopamine to regulate motivation, learning, and decision-making.[2,7] Overuse of harmful substances or behaviors can cause excessive dopamine depletion from the VTA, causing the desired and addictive feeling.[7,8] Over time, the brain misinterprets the greater dopamine production from substances as normal by making fewer rewards. As a result, such substances and behaviors appear less satisfying. Now, the only way an addicted person can get more dopamine is from external direct sources like the addictive drugs that are known to deteriorate their brain.[6,7] This causes withdrawal symptoms and the constant need for a substance after it is suppressed. The mesolimbic system stays regulated by homeostasis, balancing motivation, reward, stress, and other physiological needs.[2] Genes like DRD2 (dopamine receptors) and GABRA1 (GABA receptors) affect how these signals work, making sure rewards and motivation adjust to the body's needs. If homeostasis is disrupted, genetic differences can change how these neurotransmitters function, leading to problems with impulse control, stress, or addiction. For example, when someone is hungry, their mesolimbic system will trigger dopamine levels to rise, promoting food-seeking behavior and making the idea of eating more appealing. Dopamine and other neurotransmitter signaling play a crucial role in each of these functions, ensuring motivation, learning, and behavior are properly maintained.[2,7]

Acetylcholine Network:

The acetylcholine network is a complex system that plays a crucial role in modulating neuronal excitability, synaptic transmission, and network dynamics in the brain. Acetylcholine (ACh) modulates neuronal excitability through the inhibition of potassium currents and the activation of specific muscarinic and nicotinic receptors on cell membranes, leading to enhanced synaptic communication and dynamic network modulation.[9] The acetylcholine network helps in thinking, learning, memory, and focus by improving communication between brain cells. Improved communication allows ACh signaling to be faster, stronger, and less susceptible to distractions. This makes cognitive functions stronger and more efficient.[9,10] ACh is capable of stimulating dopamine release, which strengthens substance craving. In nicotine addiction, nicotine directly binds to nicotinic acetylcholine receptors, mimicking ACh. This overly stimulates the reward system, leading it to release abundant amounts of dopamine. Eventually, the brain is more susceptible to triggers, cravings, and relapse.[5] ACh excites neurons to promote memory, but controls inhibition to prevent any overstimulation. It regulates the amount of dopamine and serotonin to balance mood and behavior and synaptic plasticity for neural communication. Specific genes are used for their purposes in ACh networks. For example, CHAT (choline acetyltransferase) produces ACh, ACHE (acetylcholinesterase) codes an enzyme that breaks down ACh, transporter genes that create proteins to move ACh, and receptor subunits that improve synaptic transmission.[9,10] In all of these functions, neural plasticity is regulated. Neural plasticity, regulated by ACh, is essential for cognition, addiction, and homeostasis by ensuring neurons adapt efficiently without becoming unstable.

CREB-Dependent Gene Regulatory Network:

The CREB-Dependent Gene Regulatory Network is a system of genes where the transcription factor CREB (cAMP response element-binding protein) regulates multiple gene expressions, specifically by binding to cAMP. The main neurotransmitter for this network is glutamate.[1,3] Glutamine controls CREB activation by producing glutamate, exciting neurons for learning and memory, and GABA, which calms them to prevent overstimulation.[1,3,5] In cognition, CREB is essential for memory formation and learning. Neurotransmitters like glutamate, acetylcholine, and dopamine activate CREB, which strengthens neural connections and supports synaptic plasticity.[2] Activation of CREB can turn short-term memory into long-term memory. Substance abuse overactivates CREB, leading to changes in gene expression, promoting further addictive behavior. When the use of a drug stops, CREB levels also drop, creating withdrawal and cravings.[4,6] In homeostasis, CREB helps maintain neural stability by regulating responses through neurotransmitters like serotonin, acetylcholine, and GABA.[1,9,11] It can adjust neuronal activity to prevent overstimulation or underactivity, creating proper cognitive and emotional balance.

NF-κB pathways:

NF-κB is a group of transcription factor protein complexes. They play a major role in neurotransmitter signaling, synaptic plasticity, and neuroinflammation. NF-κB pathways regulate inflammation in cells and act as a master switch in response to stimuli, specifically stress or infections. In cognition, activated NF-κB helps neurons adapt by monitoring genes in learning, memory, and synaptic plasticity. Whereas in addiction, NF-κB is influenced by dopamine and glutamate signaling in the brain's reward system. When exposed to addictive substances, dopamine release triggers NF-κB activation, leading to chang-

es in gene expression that reinforce drug-seeking behavior and cravings. Over time, NF-κB strengthens reward-related pathways, making addiction more persistent. Additionally, chronic drug use increases neuroinflammation, which further dysregulates NF-κB activity, contributing to withdrawal symptoms, anxiety, and relapse. Drugs like opioids, nicotine, and cocaine are known to over-activate NF-κB, leading to long-term alterations in brain plasticity. NF-κB helps maintain homeostasis by regulating the brain's response to stress, inflammation, and immune signals by adjusting neuronal activity. Chronic stress may cause persistent NF-κB activation, which disrupts mood regulation and can cause neurodegeneration.[6]

Biological processes related to NF-κB pathways:

Because there are thousands of genes in humans, categorizing each one into a specific biological process is time-consuming and cumbersome. However, with new technological advances in biological and computational fields, we can break down the problem into simpler steps to understand which genes play a statistically significant role in certain biological processes in the human body.

The motivation behind this integrated analysis is to gain a comprehensive understanding of how NF-κB regulates gene expression across multiple biological domains. Advances in the study of NF-κB-related gene regulatory networks can help us pave better paths to fight these conditions. NF-κB is not limited to one biological process. It plays critical roles in inflammation, immune responses, cell survival, and stress signaling. By studying NF-κB, researchers can gain insights into mechanisms that affect multiple domains such as addiction (through neuroinflammation), cognition (via neural plasticity), and systemic homeostasis (through stress and metabolic regulation). We can elucidate the direct targets of NF-κB, map the complex interplay of regulatory interactions, and ultimately link these molecular insights to broader physiological and pathological outcomes. This holistic approach not only deepens our basic understanding of gene regulatory networks but also paves the path for innovative therapeutic strategies.[6]

One such study has been carried out below, where we relate the relevant genes with the NF-κB subunits and the biological processes they enrich:

Deriving the gene names interacting with each of the NF-κB subunits:

We will first find the genes interacting with each of the NF-κB subunits. We will then look for their influence on the biological processes later. The table below (Table 1) shows the sets of genes that were found interacting with each of the 5 subunits of NF-κB

Table 1: Genes that were found interacting with the specific subunits of NF-κB.

Genes interacting with **Nuclear factor NF-kappa-B p105 subunit:** ABCC2, CHUK, COPB2, CTNNB1, ESR1, HDAC1, HIF1AN, HTT, IKBKB, MAP3K8, MEN1, NFKB2, NFKBIA, NFKBIB, NOTCH1, PDCD11, PELP1, PLD3, RELA, RELB, RPS3, TNIP1, TNIP2, UL42
Genes interacting with **Nuclear factor NF-kappa-B p100 subunit:** MAP3K8, MEN1, NFKB1, NFKBIA, NFKBIB, REL, RELB, SUMO1
Genes interacting with **Transcription factor p65:** AATF, BANF1, BRD4, Brd4, CASP6, CCK, CDK5RAP2, CDK5RAP3, CHAT, CHUK, COMMD1, CREBBP, CRIP2, CTNNB1, DAXX, DHX9, EHMT1, ESR1, FGFR3, FUS, GAMMAHV.ORF73, GFI1, GSN, HDAC1, IER3, IKBKB, IRF5, KDM2A, KEAP1, KLF6, LAMP2, LMO2, MACROD1, MAPK10, MEN1, Myocd, NFE2L2, NFKB1, NFKBIA, NFKBIB, NFKBIE, nleC, NR4A1, NSD1, OGT, OTT_0753, OTT_1912, P/V, PDCD4, PIAS4, PPP2CA, PPP2R1A, PRPF40A, RAN, REL, RELB, RPS3, sctL, SETD7, SIRT1, SIRT2, SIRT6, SIX2, SNAI1, SOCS1, STAT3, TARDBP, TCF4, TGM2, TP53BP2, UBC, UBQLN1, UXT, UL42, yerA, YPO2940, YPO3877
Genes interacting with **Proto-oncogene c-Rel:** A1CF, ACOT12, ACOT8, AGPAT4, AIRIM, AKT2, ALOX5, AQP1, ARFIP2, ARIH2, ARL16, ARMC7, ASAP3, ASMTL, ATG9A, ATP6V0D2, ATP6V1C2, ATPAF2, ATXN1, BANP, BARHL2, BBS4, BCL2L15, BCL6, BID, BMF, C11orf68, C14orf119, C1orf50, C1orf74, C6orf142, C9orf72, CABP5, CACNA1S, CCNC, CCNJL, CDK18, CDKN1A, CDKN2C, CDKN2D, CDR2L, CENPX, CEP19, CFAP206, CGGBP1, CHAF1A, CHCHD2, CHIC2, CIB3, CINP, CKS1B, CLIP1, COL20A1, COL8A1, COMMD1, CPNE2, CPNE7, CPSF1, CSNK2A1, CTNNA3, CTNNBIP1, CYB5R2, DDX6, DEF6, DHPS, DMRT3, DNTT, DYNC1LI1, EFHC1, EGLN3, EHHADH, EIF3A, EIF3D, EIF4A3, EIF4E2, EIF4EBP1, EIF5A, EIF5A2, EMD, EML2, EPM2AIP1, ESRRA, ESRRG, EXOSC1, EXOSC5, EXOSC8, FAAP20, FAM200C, FAM90A1, FHL2, FKBP1B, FLAD1, FLNA, FNDC11, FNDC3B, FOXO4, FUT11, GADD45G, GLYCTK, GPKOW, GRAP, GRB2, HAT1, HDAC7, HIP1, HLA-DOA, HNRNPF, HSD17B14, HSPB7, IL36RN, KCTD6, KCTD7, KIAA0100, KLHL32, KLHL42, KRTAP9-4, L3MBTL2, LASP1, LENG1, LGALS14, LMO1, LMO2, LNPEP, LSM2, MAD2L2, MAGOHB, MAP3K7CL, MAPRE3, MB21D2, MBD3, MEMO1, MENT, MID2, MIEN1, MIPOL1, MMP2, MORF4L1, MORF4L2, MRPL10, MSRB3, MTURN, MYO15B, NABP1, NAGK, NCK2, NDUFAF3, NECAP1, NEDD9, NEIL2, NEUROG3, NFE2L2, NFKB1, NFKB2, NFKBIA, NFKBIB, NFKBIE, NGB, NME7, NMRK1, NOL9, NOXA1, NR2C2AP, NRSF, NUDT10, NUDT14, NXT2, OAZ3, OSGIN1, OSTF1, OTUB2, OTUD4, OTT_0753, OTT_1912, PADI3, PAK1IP1, PARVG, PATE1, PCYT2, PELI2, PGAP6, PICK1, PIH1D2, PLCB4, PLEKHN1, PM20D2, POLE2, POLR1A, POLR2L, POP5, PRDM10, PRKAA2, PRKAB2, PRPF31, PRPF40A, PSMA1, PSMA4, PSMB8, PTGER3, PTK6, QARS1, R3HDM2, RAB29, RAB41, RABIF, RAD23A, RAD51D, RASSF5, RBM39, RECK, RELA, RELB, REXO1L1P, RIPPLY1, RNF213, RNF6, RPS3, RSPH14, RTL8C, RUNX1T1, S100A1, SAMD11, SAT1, SCNM1, SDCBP, SEC14L4, SEC31A, SEPTIN7, SHC3, SIX1, SLC39A13, SLC41A3, SLC6A12, SMG9, SNRNP25, SOCS1, SPATC1L, SPG21, SPRY1, STAMBPL1, STAT3, STK16, SULT2B1, SUOX, SZT2, TBC1D21, TBC1D22B,
TCL1A, TCP11L1, THOC1, TLE5, TLX3, TNIP2, TNS2, TRAPPC2L, TRAPPC6A, TRIM68, TRIM74, TSC22D3, TSEN15, TSSC4, TSSK3, TTC19, TTC21A, TXNL4B, UBASH3A, UBASH3B, UBE2D4, UBE2K, UBE2Z, UBXN7, UBC, UXT, UL42, VPS25, WNK3, YARS, YJU2, YPO2940, YPO3877, ZBTB16, ZDHHC24, ZIC1, ZNF417, ZNF446, ZNF550, ZNF564, ZNF572, ZNF688, ZNF76, ZNF765
Genes interacting with **Transcription factor RelB:** COMMD1, GSK3B, NFKB1, NFKB2, RELA, SMARCC1

Deriving the biological processes enriched by each set of interacting genes:

We will now find the biological processes enriched by these genes, one set at a time. Only those biological processes are selected where the role of these genes has a higher significance (i.e., p-value < 0.05). This means that the influence of those genes on those biological processes is statistically significant, provided the sample size of the study done for those processes is high enough. Out of these biological processes, the top 10 processes are printed against the scale of -log10(p-value). This negative log (with base 10) keeps the values within the graph and helps with visualization since the real p-values are too far away in the linear scale. Here are the top 10 biological processes for each set of genes mentioned in Table 1. These are shown in Figures 1 through 5. Figure 1 below shows the top 10 biological processes that were influenced by the genes mentioned in Row 1 of Table 1 above. These are the genes that interact with subunit 1 of NF-κB.

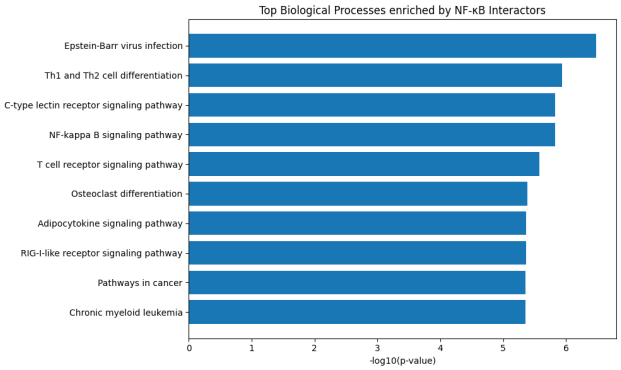


Figure 1: Top 10 biological processes enriched by genes that interact with subunit 1 (NF-κB B1-p50).

DOI: 10.36838/v7i10.108

Figure 2 below shows the top 10 biological processes that were influenced by the genes mentioned in Row 2 of Table 1 above. These are the genes that interact with subunit 2 of NF-κB.

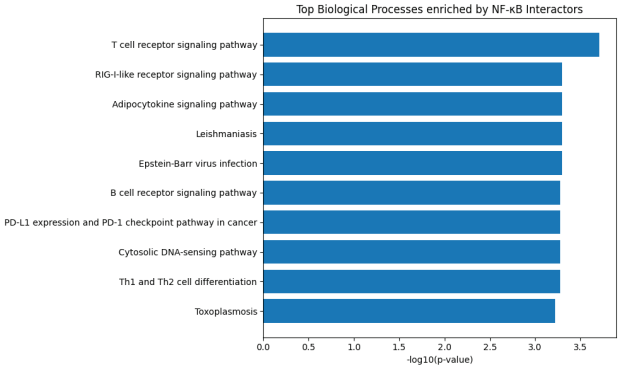


Figure 2: Top 10 biological processes enriched by genes that interact with subunit 2 (NF-κB2-p52).

Figure 3 below shows the top 10 biological processes that were influenced by the genes mentioned in Row 3 of Table 1 above. These are the genes that interact with subunit 3 of NF-κB.

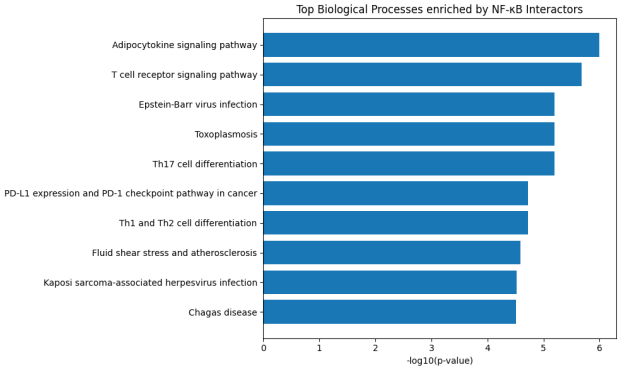


Figure 3: Top 10 biological processes enriched by genes that interact with subunit 3 (RelA-p65).

Figure 4 below shows the top 10 biological processes that were influenced by the genes mentioned in Row 4 of Table 1 above. These are the genes that interact with subunit 4 of NF-κB.

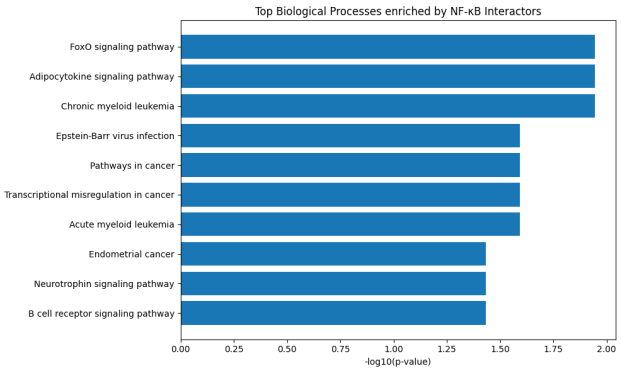


Figure 4: Top 10 biological processes enriched by genes that interact with subunit 4 (cRel).

Figure 5 below shows the top 10 biological processes that were influenced by the genes mentioned in Row 5 of Table 1 above. These are the genes that interact with subunit 5 of NF-κB.

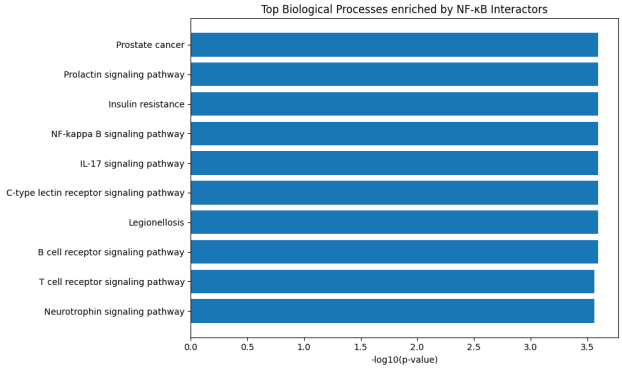


Figure 5: Top 10 biological processes enriched by genes that interact with subunit 5 (RelB).

Assessing gene commonality across the subunits:

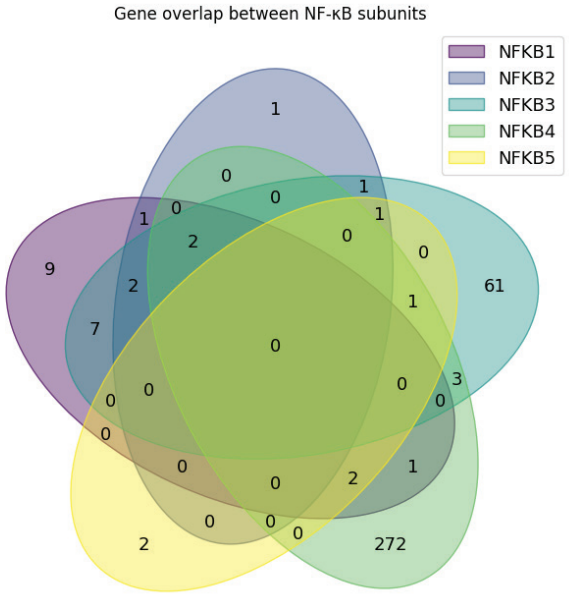


Figure 6: Assessing gene overlaps across the 5 subunits in NF-κB.

The Venn diagram in Figure 6 shows the overlap of the distinct genes interacting with each of the subunits of NF-κB. It is noteworthy that no single gene is common across all 5 NF-κB subunits. The data also goes on to show that each subunit is substantially exclusive and represents a diverse set of genes. Biological processes that are common to such a diverse set of genes can show promising relationships among the NF-κB subunits and genes influencing those.

Deriving common biological processes across the 5 sets of genes:

In this step, we will find the biological processes common across the 5 sets found above. Instead of taking just the top 10 biological processes, we take the entire record set of each of the 5 sets of biological processes formed above and carry out an intersection among them to find the common ones. The

function created for this purpose allows for choosing n number of records from each set of the biological processes and finding the overlap based on those. As an example, 100 records were chosen from each of the 5 record sets, and 9 biological processes were found to be common among them as of this writing. These are shown in Table 2 below. Note that this set can change if the program is run sometime in the future when the underlying APIs return different data.

Table 2: Biological process enhanced by all 5 sets of genes that interact with each of the NF-κB subunits.

Adipocytokine signaling pathway
Epstein-Barr virus infection
T cell receptor signaling pathway
Neurotrophin signaling pathway
Transcriptional misregulation in cancer
B cell receptor signaling pathway
PD-L1 expression and PD-1 checkpoint pathway in cancer
Chronic myeloid leukemia
Small cell lung cancer

Discussion

In this study, I have compared the attributes of the four genetic networks discussed above. The mesolimbic system helps the brain with dopamine levels and regulation. Correct amounts of dopamine at the correct times can increase motivation and learning to improve cognition. However, addiction can alter dopamine signaling, creating feelings of craving and the sensation to always want more.[7] Altered dopamine signaling makes it harder to balance stress and motivation, which can compromise neuronal homeostasis. The Acetylcholine network modulates neuronal excitability and how well the brain can react to stimuli.[9] The more connected the neurons are, the faster they can communicate. Faster communication creates faster reaction times and improves levels of cognition. Dopamine levels influence excitability levels of neurons, and repeated substance exposure affects the excitability in different ways, specific to the substance.[5] Maintaining neuronal excitability properly preserves neuronal homeostasis. The CREB-dependent network is important in coordination and communication between cells.[1] Faster communication within cells improves cognition and reaction time. Addiction can overstimulate reward pathways and can create tolerance from the repetition of substance abuse.[4] The network maintains neural stability. NF-κB pathways regulate neuroplasticity and play a crucial role in immune responses. They help neurons adapt to stimuli, which improves cognition. Addiction can create excessive dopamine release that can trigger chronic responses and neuroinflammation. They regulate stress and immune response.[6]

All 4 networks ultimately influence gene expression and contribute to various aspects of brain function, including reward processing, attention, memory, learning, and neuronal plasticity. As far as differences go, the mesolimbic system and the Acetylcholine network operate at the intercellular level, and their effects are often rapid and transient. CREB-dependent networks and NF-κB pathways are at the subcellular level, and they regulate gene expression for hours or even days.

We have found the biological process enhanced by the genes that interact with the NF-κB subunits. This finding can influence the direction in which those biological processes are being studied. For example, the biological process found to be common across the 5 NF-κB subunits suggests a stronger correlation of the entire NF-κB pathway with the life cycle of those biological processes. The experts studying NF-κB can provide valuable input into the research of these biological processes.

Further research should explore the temporal dynamics of gene expression through different addiction stages. Such research can help pinpoint the roles of each genetic network at every stage, which can be used for more specific and personalized treatments. For example, we can differentiate which processes create homeostatic responses and which are used in pathological states in severe addiction. Researching further into genetic polymorphisms and mutations can help personalize treatment as well.

Gene expression cannot always reflect actual protein use. For example, post-transcriptional regulation, protein modification, and degradation all affect protein quantities, and the functional significance remains as inferred.[6] The data used to determine NF-κB interactions are from network-based sources and do not include behavioral data.

This study supports that genetic networks are responsive to cognitive, addictive, and homeostatic demands, often bidirectionally.[6] The data analysis showed that genetic networks can influence a broad range of biological pathways throughout the body.

Conclusion

The Mesolimbic System and Acetylcholine Network interact closely in regulating reward, learning, and motivation.[7,9] Dopamine, the key neurotransmitter in the mesolimbic system, is influenced by acetylcholine, the key neurotransmitter in the acetylcholine network, which can enhance or inhibit dopamine release.[5,9] The NF-κB pathway is known for its role in immune response and interacts with the brain's neurotransmitter systems, usually under conditions of stress or neuroinflammation. This can disrupt dopamine signaling in the mesolimbic system, leading to depressive or cognitive-deficiency symptoms.[6] All four systems are connected because neurotransmitters (dopamine and acetylcholine) affect gene expression as seen in CREB and NF-κB pathways, which in turn change how neurons function over time.[1,6]

Deriving the genes interacting with each of the subunits of the NF-κB pathway and finding the common biological processes influenced by those genes shows that we should look at the NF-κB subunits in totality while also assessing those common biological processes. It also asserts the importance of considering the interplay between those genes while studying the proteins in the relevant NF-κB subunits and the biological processes influenced by them.

Looking forward, there is still much to be uncovered on genetic networks because of their quantity and specificity. For example, the total number of genetic networks or how many of them exist are questions that are still unanswered. We are

DOI: 10.36838/v7i10.108

also still unaware of the comprehensive locations of genetic networks and how far they span. Some factors of networks to research further include the longevity of networks, genetic variants in networks, the usage of these systems to create therapeutics for diseases, and the impact of external factors on them.

■ Acknowledgments

I attest that the ideas, graphics, and writing in this paper are entirely my own. I would like to thank my mentor, Dr. Jorge Avila.

■ References

1. Kida, S. A Functional Role for CREB as a Positive Regulator of Memory Formation and LTP. *Exp. Neurobiol*. **2012**, *21* (4), 136.
2. Hyman, S. E. Neurotransmitters. *Curr. Biol*. **2005**, *15* (5), R154–R158.
3. Sakamoto, K.; Karelina, K.; Obrietan, K. CREB: A Multifaceted Regulator of Neuronal Plasticity and Protection. *J. Neurochem*. **2011**, *116* (1), 1–9.
4. Carlezon, W. A.; Duman, R. S.; Nestler, E. J. The Many Faces of CREB. *Trends Neurosci*. **2005**, *28* (8), 436–445.
5. Wittenberg, R. E.; Wolfman, S. L.; De Biasi, M.; Dani, J. A. Nicotinic Acetylcholine Receptors and Nicotine Addiction: A Brief Introduction. *Neuropharmacology* **2020**, *177*, 108256.
6. Nennig, S.; Schank, J. The Role of NFkB in Drug Addiction: Beyond Inflammation. *Alcohol Alcohol* **2017**, *52* (2), 172–179.
7. Serafini, R. A.; Pryce, K. D.; Zachariou, V. The Mesolimbic Dopamine System in Chronic Pain and Associated Affective Comorbidities. *Biol. Psychiatry* **2020**, *87* (1), 64–73.
8. Trainor, B. C. Stress Responses and the Mesolimbic Dopamine System: Social Contexts and Sex Differences. Horm. Behav. **2011**, *60* (5), 457–469.
9. Roach, J. P.; Eniwaye, B.; Booth, V.; Sander, L. M.; Zochowski, M. R. Acetylcholine Mediates Dynamic Switching between Information Coding Schemes in Neuronal Networks. *Front. Syst. Neurosci*. **2019**, *13*, 64.
10. Riese, H.; Munoz, L. M.; Hartman, C. A.; Ding, X.; Su, S.; Oldehinkel, A. J.; Van Roon, A. M.; Van Der Most, P. J.; Lefrandt, J.; Gansevoort, R. T. Identifying Genetic Variants for Heart Rate Variability in the Acetylcholine Pathway. *PLoS One* **2014**, *9* (11), e112476.
11. Wen, A. Y.; Sakamoto, K. M.; Miller, L. S. The Role of the Transcription Factor CREB in Immune Function. *J. Immunol*. **2010**, *185* (11), 6413–6419.
12. Tanis, K. Q.; Duman, R. S.; Newton, S. S. CREB Binding and Activity in Brain: Regional Specificity and Induction by Electroconvulsive Seizure. *Biol. Psychiatry* **2008**, *63* (7), 710–720.

■ Author

Riya Shukla is a freshman at West Chester East High School in West Chester, Pennsylvania, USA. Riya has a passion for studying medicine and understanding human anatomy. She is looking forward to making an impact in the field of health care.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

The Form and Function of Chromatin Organization

Ryota Watanabe

Crimson Global Academy, Level 3, 18 Stanley St, CBD, Auckland 1010, New Zealand; ryotawatanabe1972@gmail.com

ABSTRACT: The structure of chromatin and how it is used for cellular function is fundamental to multicellular organisms. The form and function of chromatin organization affect almost every aspect of gene expression, cell division, and various other cellular processes. This review article aims to provide a comprehensive view of the intricate structural organization of the chromatin and its functional relevance. Beginning with an introduction to DNA and heterochromatin, then focusing on the individual nucleosome and its components, the core histones H2A, H2B, H3, and H4. Subsequently, the article covers chromatin architecture, including its past model of 30 nm hierarchical folding and the current consensus of irregular fibers, as well as how the chromosome is structured around a protein scaffold. Next, the chromatin remodeling complexes, their four categories and distinct functions, and how chromatin is manipulated during mitosis are discussed. Ultimately, I explore how the form of chromatin affects gene expression through physical properties such as liquid-liquid phase separation and nuclear compartmentalization.

KEYWORDS: Molecular and Cellular Biology, Genetics, Chromatin Packaging, Chromatin Compaction, Chromatin Condensation.

Introduction

The diploid cells, with two sets of chromosomes, in a human body contain 6 billion base pairs (bp) of DNA (Deoxyribonucleic acid), which approximately equals 2 meters when stretched. This is packaged in a minuscule nucleus that is 10 micrometers in diameter. Despite this significant compaction, the cell can interpret DNA and the state that it is in to carry out vital processes like cell division and gene expression reliably and accurately. Understanding how chromosomes are organized is an indispensable field of research, as evidenced by the scientific attention and studies conducted about it in the past. It began in the middle of the 19th century when Gregor Mendel postulated that traits were inherited as discrete units without referring to the chemical or structural organization of what is inside the "discrete units". Afterwards, researchers would begin to investigate the basis of heredity, giving rise to the scientific discipline of genetics. Subsequently, Walther Flemming used dyes and basic microscopy to discover the unit of genetic material and named it "chromatin" in 1879.[1] Following that, in 1891, Hermann Henking discovered the X-shaped chromosome while studying spermatocyte divisions of a firebug.[2]

Further insights on DNA arose when the electron micrograph images of the DNA forming "beads on a string" were published in 1974.[3] We now understand that the reason for this appearance is due to the DNA being spooled into the nucleosome. Similarly, there have been massive recent innovations in assays and technologies, such as X-ray scattering, partial decondensation of chromosomes, and ChromEMT (ChromEM tomography, an electron-microscopy staining technique that marks nuclear DNA without changing its structure, permitting better visualization of 3D chromatin conformation). These foundational studies have significantly enhanced our understanding of the structure of chromatin.

Several subsequent studies have allowed scientists to learn how the DNA sequence influences phenotypes and mutations that induce certain diseases. However, it is now known that the proteins that regulate the architecture of DNA are equally key in regulating cellular function. Mutations in genes encoding those proteins can promote conditions like cancer. This new aspect of genetics may be the path that could lead to a multitude of new discoveries and opportunities in treating cancer and chromosomal abnormalities. This could have a widespread impact due to the significant proportion of the population affected by these disorders. Specifically, there were 2,001,140 new cases of cancer in the US in 2024,[4] and an estimated global risk of 25% of getting cancer during a person's life.[5]

In this review, I explore the currently available literature to illustrate the form and function of chromatin structure and how the cell attains it. I discuss how fundamental functions such as cell division and gene expression rely on chromatin architecture and present how misregulation of chromatin can be a catastrophic event for many cells. I specifically focus on how misregulation of key components of chromatin organization leads to damage in cells, which leads to certain disorders such as cohesinopathies and cancer. Understanding the detailed activities of components of chromatin organization and their relevance to the disorders will assist the development of novel treatments by offering potential drug targets.

DNA, Hetero, and Euchromatin:

DNA is the essence of all living organisms on Earth, which encodes all proteins used by cells for every function. It is made of nucleotides that consist of a phosphate group, five carbon sugars, and a nitrogenous base. While all DNA is composed of the same molecules, the unique arrangements of the bases adenine (A), thymine (T), cytosine (C), and guanine (G) store the coding information for every gene and protein.

DOI: 10.36838/v7i10.114

The complementary base pairing between adenine and thymine, and between cytosine and guanine, connects two DNA strands that are wrapped around each other in a double helix form, with approximately 10 bp per turn, where it assumes its most stable form.[6] However, as the bond between the sugar and the base is asymmetrical, as the angle formed by the glycosidic bond between the base and sugar is not aligned at exactly 90 degrees,[7] the spaces between backbones form two grooves: the major groove and the minor groove. The major groove is wider than the minor groove. The existence of these grooves is key for proteins to bind and recognize DNA sequences outside the DNA (Figure 1). However, DNA by itself, as an individual entity, cannot contribute to the proper function of the cell.

When DNA is bound to proteins and RNA, chromatin, the building unit of chromosomes, is formed. Chemical interactions with the DNA can only occur if the DNA is physically accessible. Its openness is determined by many factors that will be explained later in this review, such as nucleosome positions, histone modifications, transcription, and chromatin compaction.[8] There are two broad categories for chromatin: hetero or euchromatin. Heterochromatin describes an inaccessible state where the genes in DNA are repressed, while the latter is accessible and has active genes. As such, heterochromatin is much more DNA and nucleosome dense than euchromatin,[9,10] and the diffusion in heterochromatin is also less efficient compared to euchromatin.[11]

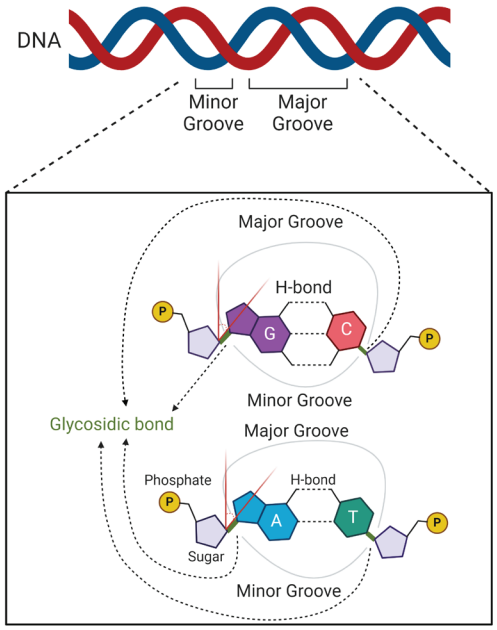


Figure 1: DNA major and minor groove formation. Schematic illustration of groove formation in AT and CG base pairs due to asymmetry in glycosidic torsion angle not equaling 90 degrees (in red line angle). The glycosidic torsion angle is the angle between the phosphate backbone and the glycosidic bond attaching it to the nitrogenous base. This asymmetry is what allows proteins and transcription factors to recognize and bind desired DNA regions.

However, the categorization of chromatin is not a recent development. The nomenclature of heterochromatin and euchromatin originates from Emil Heitz in 1928, when he published the paper 'Das heterochromatin der moose'. He used it to describe the differences detectable by appropriate chromosomal stains. Specifically, euchromatin becomes invisible during late telophase (the final stage of mitosis), while heterochromatin is heteropycnotic, taking up more stain or being more tightly coiled after telophase.[12]

How Does a Cell Package All of its DNA into a Smaller Nucleus? Nucleosomes and Histones:

Since each haploid human cell contains 3 billion bp, each being 0.34nm long,[13] in each diploid cell, there would be 2 meters of DNA. How do cells package all of this DNA into microscopic nuclei with diameters of 5-20μm? A large length compaction of over 100,000-fold. Cells compact the DNA into 'beads on a string' through nucleosomes - the length of DNA wrapped around a histone octamer- wrapping the DNA around the core of a histone octamer (Figure 2). Histones are positively charged proteins that facilitate electrostatic interactions with the negatively charged backbone of DNA to wrap it around themselves, leading to nucleosome formation. There are four core types of histones in humans: H2A, H2B, H3, H4, and a linker histone H1.[14] The four core histones share properties such as their C-terminal domains, which provide histone-histone interactions to form the column-like octamer,[15] which takes up most of their mass.

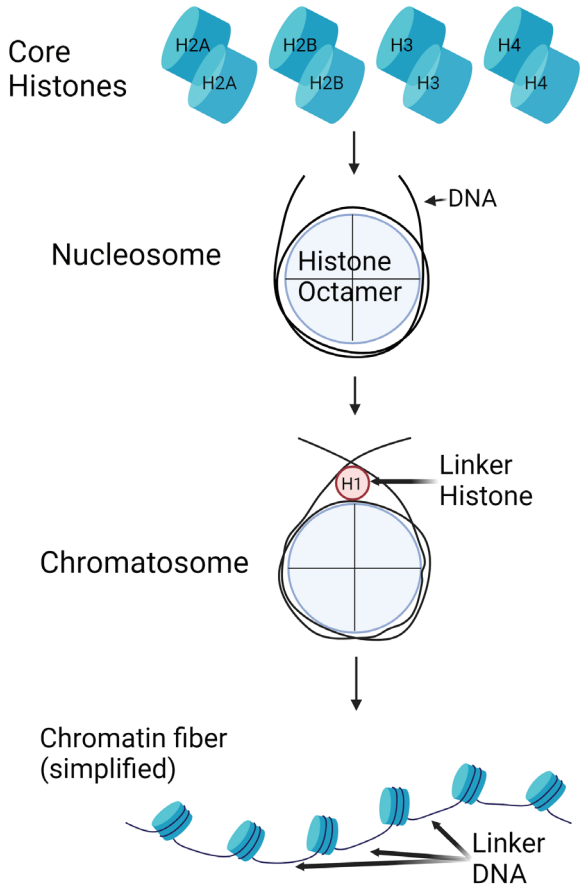


Figure 2: The making of chromatin fiber. Schematic illustration of the process histones go through to become part of a chromatin fiber. With the help of proteins such as assembly remodelers, the core histones form a histone octamer, around which DNA is subsequently coiled, forming a nucleosome. The nucleosome, with the addition of a linker histone, turns into a chromatosome, and the DNA now goes through 2 full turns instead of 1 and ⅔. Many nucleosomes form the chromatin fiber. This makes the DNA very tightly compacted in a regulated manner.

The four histones H2A, H2B, H3, and H4 are called canonical proteins because they are the standard of their respective types, recognized as the norm. However, there are a multitude of variants, such as H3.3, that are non-canonical as they differ from this norm, but I will not discuss them further here as it warrants a focus as a separate topic.[8] Their N-terminal tail provides strong contacts to the DNA that the nucleosomes bind to.[16] The N-terminal tails are readily available for post-translational modifications to be catalyzed by enzymes, playing an important role in epigenetic signaling.[15] They are also in place to conduct inter-nucleosomal interactions in condensed chromatin structures, which help organize higher-order chromatin structures.[17]

Table 1: Glossary of terms.

Term	Definition
Chromosome	A cellular entity containing DNA and protein molecule that is packed into a more complex structure.
Chromatin	The complex of proteins, RNA, and DNA that forms the state of the genome.
Nucleosome	The repeating subunit of chromatin inside the nucleus.
Linker DNA	The length of DNA that connects adjacent nucleosomes (Figure 2).
Chromatosome	The unit of chromatin includes the linker DNA, the nucleosome and the linker histone H1.
Chromatid	Two individual chromosomes that are joined at the centromere that are newly replicated.
Solenoid	A coil in the form of a long cylinder (Figure 3).
Scaffold	A protein structure that provides support for the chromosome. (Figure 3)
Octamer	A complex made of 8 subunits of histones (Figure 2).
ATP (Adenosine Triphosphate)	The universal chemical source of energy in the cell.
ATPases	Enzymes that Proteins that convert ATP to ADP (Adenosine Diphosphate)
Transcription	Synthesis of messenger RNA (mRNA) from DNA sequence as template
Translation	Decoding of mRNA sequence. into amino acids

The product of the 4 abovementioned histones, the nucleosome is the unit of chromatin made up by the spool-like histone octamer wrapping DNA in a left-handed superhelical manner.[15] Its main role is to condense the DNA, but it also has other roles such as repressing processes involving DNA, like transcription, replication, and repair. A nucleosome is created by first gathering two copies of each protein H2A, H2B, H3, and H4 proteins to come together to form a histone octamer, which is what constitutes the nucleosome (Figure 2).[18] The octamer forms through the following steps: firstly, two of the H3:H4 dimers and H2A:H2B dimers are formed, then the two H3:H4 dimers merge through H3:H3 interactions to form a tetramer, and finally, the two H2A:H2B dimers combine with the tetramer via H4:H2B interactions.[15] Nucleosomes are known to condense DNA by around 6~7 times.[8] Specifically, they bend approximately 146 bp of DNA around the histone octamer for approximately 1 2/3 turns.[16,19,20] The core histone octamer domains fold 120 bp, while the remaining 13 bp are bound by H3 N-terminal alpha helices. These interactions help maintain the stability of the nucleosome. The folding is caused by roll-based bending of bp into the minor and major grooves that face the nucleosome.[21] The nucleosome also twists the DNA further, such that the number of bp per turn goes from 10.5 to 10.2.[16]

Although the histone proteins predominantly fold and shape the DNA organization, the sequence of the DNA affects its binding affinity to the histone octamer, creating the phenomenon of sequence-based DNA binding.[15] The affinity varies by 3 orders of magnitude depending on the DNA sequence.[22] It is most optimal for nucleosome binding when bendable bp, such as AT & TA, are at the repeated elements every 10 bp that directly interact with the nucleosome.[23] However, having many repeats of AT/TA is inhibitory to nucleosome binding, thus promoting promoter accessibility, nucleosome depletion, and transcriptional activity.[24,25]

Similarly, the DNA bound by the histone octamer frequently undergoes a process known as "DNA Breathing" that exposes the protein binding sites on the nucleosomal DNA. DNA breathing does this by transiently binding and unbinding to the histone octamer, which can facilitate sequence-specific protein binding if the binding affinity to naked DNA or the local concentration of the protein is high enough to compete with core histone binding.[18,26] Without DNA breathing, the nucleosome would protect the DNA from nuclease digestion, as well as greatly restrict the binding of regulatory factors. Even when breathing, the binding affinity of most DNA-binding factors to nucleosomal DNA is reduced by $10^{3\sim5}$ compared to naked DNA.[15] The length of DNA that connects two nucleosomes is the linker DNA. While the linker DNA length varies from 20-80 bp, the H1 family of proteins, the linker histones, wraps linker DNA near the center and the entry/exit points of the nucleosome, promoting further compaction and stabilization of DNA. This coils an additional 20 bp of DNA, ending up in ~2 oriented full turns around the whole chromatin and linker histone, forming the chromatosome (Figure 2).[27] The functions of linker histones include: promoting folding and assembly of higher-order chromatin structures; changing nucleosome spacing on DNA, regulating specific gene expression, and protecting linker DNA.[28–30]

Not only can histones compact the DNA, but their N and C-terminal tails can be post-translationally modified in various ways, such as acetylation, phosphorylation, methylation, SUMOylation, and ubiquitination. These alter the charge and structure of the tail and thus affect their binding to DNA, ultimately changing the condition of the chromatin and how much the genes are expressed, leading to changes in cellular events such as mitosis. However, in the interest of focus, post-translational modification of histones is not covered at length in this paper. An extensive review on this topic is published elsewhere.[31]

This knowledge about nucleosomes and their post-translational modifications can be useful as they are known to play a role in many cellular processes involving DNA, such as DNA repair and transcription. Moreover, their alteration can affect broader processes, such as development and aging, as well as result in diseases such as cancer and intellectual disability, making their understanding more important.[32]

DOI: 10.36838/v7i10.114

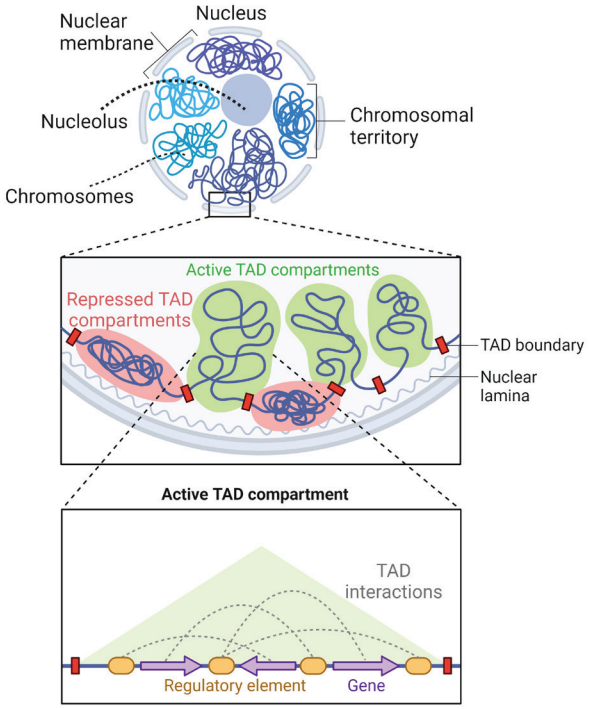


Figure 3: Organization of TADs, the constituents of chromosomes. Schematic diagram showing the nuclear structure and organization of the TADs of the chromosomes that lie close to the nuclear membrane. The TADs are composed of multiple genes and elements that physically interact with each other. The implications of 3D structuring add further complexity to the regulation of gene expression and chromatin organization in the cell.

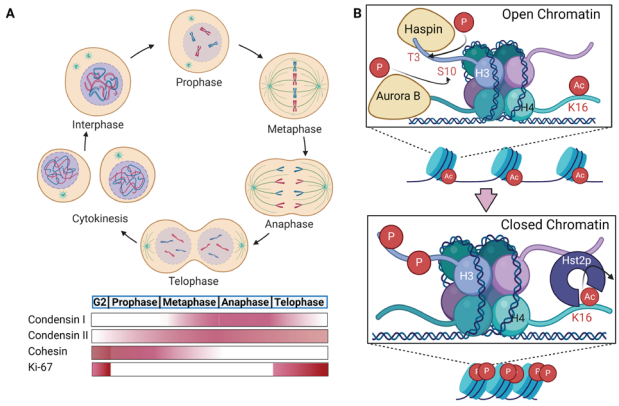


Figure 4: Mitosis and chromatin organization. **(A)** Shows the mitotic cell cycle from interphase to cytokinesis. Shows the nucleolus dissolving, and the genetic material condensing into chromosomes with sister chromatids, which are then separated into separate cells during cytokinesis. The bottom shows the mitotic phases and shows the concentrations of the proteins condensin I, condensin II, cohesin, and Ki-67 at each phase. The lighter the color, the lower the concentration. Adapted from,[37] **(B)** Schematic diagram showing the steps in which histone modifications work to condense chromatin into chromosomes in early mitosis. Haspin modifies the histone H3 by adding a phosphate group to the third threonine (T3). This triggers the phosphorylation of the 10th Serine on the histone H3 (S10) by Aurora B, which brings about the Hst2p that removes the acetyl group from the 16th lysine on the histone H4. This brings about H4 tail and acidic patch interactions with the adjacent nucleosome, thus inducing heterochromatin and chromosome formation. This mechanism demonstrates how important histones are not only in processes such as gene expression but also in mitosis.

Chromatin in Cell Division:

Mitosis is the process that cells undertake, which, in short, replicates and doubles the genetic material, then organizes and arranges it into groups so that the cell can split in half with both daughter cells having equal amounts of DNA. More specifically, the process involves five main phases: interphase, prophase, metaphase, anaphase, telophase, and cytokinesis, in that order. Interphase can be further split into the Growth/Gap (G)1, Synthesis (S), and G2 phases, in that order.

This progression through phases in mitosis is regulated by proteins that add a phosphate group (phosphorylate) to other substances from ATP, called cyclin-dependent kinases, abbreviated CDKs. For example, they can target and phosphorylate the tumor suppressor protein retinoblastoma.[33] CDKs require proteins called cyclins to activate them by binding to them and phosphorylating a certain part of them. Each class of cyclins corresponds to the phase at which they activate CDKs, like the G1-phase cyclins, G1/S-phase cyclins, S-phase cyclins, and M-phase cyclins. The last one is the cyclin, which drives the start of mitosis. There are 20 CDKs and 29 cyclins in humans.[34,35] Early in mitosis, the series of phosphorylations, H3T3 by Haspin, then H3S10 by Aurora B, causes recruitment of a lysine deacetylase Hst2p. It removes an acetyl group from H4K16, a modification that was previously interrupting internucleosomal condensation, to promote chromatin condensation (Figure 4).[36] When chromosomes are most compacted and segregated (anaphase), they are massively condensed longitudinally.

DNA can also be compacted in other ways: while the DNA is simply replicated by DNA polymerases on the irregular fiber of nucleosomes during interphase, during early prophase, the fibers begin to condense into loops around a central chromosome scaffold, as will be covered later in this paper. The condensed chromosomes are segregated by mitotic spindles. In late anaphase and telophase, the chromosome is then decondensed back into nucleosome fibers into its interphase state. However, the mechanism of this decompaction is different from just reversing compaction.

Although entry into mitosis is largely regulated by kinases, to decondense, not only are the relevant kinases deactivated, but the mitotic phosphorylation is also reverted. The phosphatases, removers of phosphates using water, PP1 and PP2A, are the main proteins in this process. PP1γ dephosphorylates the histone H3 at several points after being brought to anaphase chromosomes by the recruiting subunit Repo-Man or by another protein, Ki-67. Meanwhile, PP2A is implied to ensure the timely exit from mitosis through dephosphorylation.[38] In addition to dephosphorylation, Aurora B kinase must be evicted as well. The group of proteins AAA+-ATPase p97 and the cofactors UFD1 and NPL4 guarantee Aurora B's removal. It is implied that they use the ubiquitin proteasome pathway. It is a way that the cell warrants degradation of proteins by attaching a recognizable marker, ubiquitin, to them such that it can be transported to a proteasome for its destruction, as AAA+-ATPase p97 is a system that typically recognizes ubiquitin on tagged proteins.[37] This then implies that Aurora B is ubiquitinated during the later stages of mitosis. It is also speculated that the removal of phosphorylations by Aurora B is

responsible for the halting of the protein condensin's activity for decondensation.[39] Since this decondensation is required for mitotic exit to occur, the absence of factors such as PP2A and its cofactors can delay mitotic exit.[40] This delay could potentially lead to hazardous consequences, such as DNA damage and aneuploidy, a potential cause for tumorigenesis and cancer, making mitotic decondensation a vital topic for further research.

Chromatin Organization in Interphase:

In the last couple of decades, there have been many hypotheses on how the chromatin is compacted. However, there have been more prominent models, such that recently, chromatin has been thought to be mostly compacted by either one of two models, the hierarchical folding model or the radial loop model. (Figure 5) The hierarchical folding model proposes that the chromatin is compacted into a large-scale 30-nm diameter fiber.[41] This is thought to occur by folding into a solenoid or zig-zag shape.[42] These fibers are then progressively used to form higher-order chromatin structures of larger diameters, involving ~100 then ~200 nm, to finally become large interphase chromatin fibers. They may later be anchored to protein scaffolds to form condensed mitotic chromosomes.[8] But of the two models, the current consensus is that the majority of cells use the radial loop model. This is because most of the evidence for the hierarchical model comes from *experiments in vitro*. This is where the artificial chromatin has an array of very regularly spaced nucleosome binding sites with uniform linker DNA length. Unfortunately, there are very rarely regions with these conditions in vivo (meaning inside a living organism), because typically cellular chromatin is found to have variable linker DNA length and histone post-translational modifications, which is a significant critique of this model. There are some instances, such as in some genomic regions and mostly inactive chromatin, like in avian erythrocytes, as well as in regions of very high levels of linker histones. There have also been more data against this model in recent years, and as such will not be discussed further in this paper. [8,9,19,43–45]

The other model, the irregular model, is the current consensus that proposes that the mitotic chromosomes are condensed in irregularly arranged forms, leading to a fractal nature. The decreased physical constraint permits a more dynamic and flexible organization than its static counterpart, the 30 nm diameter fiber.[46,47] Specifically, chromatin in this model is formed into irregular, yet dynamic 10 nm-diameter fibers. The accessibility of the fibers according to computer simulations could come about as a result of nucleosomal fluctuations exposing the genomic DNA and facilitating the mobility of diffusing proteins.[46] The 10nm nucleosomal irregular fibers of chromatin are thought to remain irregular during interphase, but there have been models proposing higher orders of irregular fibers, specifically forming many condensed domains out of these fibers, so that they look like "chromatin liquid drops".[48,49] These drops of chromatin are thought to be formed by the macromolecular crowding effect and specific proteins like cohesin and CTCF (CCCTC binding factor),[50–53] and/or condensin II.[39,54]

Additionally, the physical packaging units of DNA are known as TADs (Topologically Associated Domains). Interestingly, the genes within the same TADs actively interact with each other often, but not with those in other TADs. These TADs correspond to LADs (Lamina Associated chromatin Domains),[55] which are domains that are associated with nuclear lamina - a structure near the inner nuclear membrane, composed of fibrous proteins known as lamins.[56] These chromatin domains serve to help change timing during cell differentiation and transcriptional regulation, as well as for other purposes. Enhancer-promoter interactions are likely limited to regions within the same TAD, as seen in Figure 3.[30]

Nucleosome fibers are arranged into mitotic chromosomes by the activity of certain proteins. This is done so that the chromosomes can be individualized and so that sister chromatids (Table 1) can be separated.[57,58] However, recent research has shown that condensin - a key protein in chromatin condensation into chromosomes – is not required for chromatin to organize into lumps in yeast.[59] This has induced proposals that there is a condensin-independent condensation which occurs in parallel with condensin-dependent condensation, where neighboring nucleosomes are attracted to each other, due to interactions between H4 tails and nucleosomal acidic patches.[36] This has been postulated to be a small-scale form of chromatin compaction which occurs inside the loops of DNA, while the condensins compact at a larger scale, looping long stretches of DNA.[38]

There are some other hypotheses, such as hyperactive condensin DNA supercoiling - the number of DNA bp per turn of the helix decreasing, determining the amount of strain on the DNA, being a contributor to mitotic compaction.[60] This is unsurprising due to how important mitotic compaction is. Therefore, it stands to reason that there would be many types of compaction, including, albeit redundant ones, as any that enhanced compaction would be positively chosen by evolution.[61]

Chromatin Condensation into Mitotic Chromosomes:

The difference in compaction of chromatin between mitotic chromosomes and interphase chromosomes is relatively subtle, where the compaction of one chromosome's worth of DNA is ~10^4 and ~$8.5*10^3$ fold, respectively.[62] This is even though estimations have predicted the difference to be 2-3 fold, which in retrospect seems like an overestimate.[63] The chromatin chain also appears to be more flexible in mitotic chromosomes, allowing for the nucleosomes to be more tightly packed.[9] Although their compaction is relatively similar, their form and arrangement are significantly different. How does the cell cause this? The cells can induce this by using 6 essential proteins that are required for mitotic condensation to occur. The only required proteins to form the chromosome *in vitro* are topoisomerase IIα, cohesin, condensin I and II, and the chromokinesin KIF4A. They also appear to be the major, if not only, essential components of the chromosome scaffold.[64,65] First found by electron microscopy, a central 'scaffold' in the shape of metaphase chromosomes organizes the looping of the DNA along

DOI: 10.36838/v7i10.114

the chromatid and is responsible for the chromosome's basic shape as seen in Figure 5.[57,61]

Specifically, condensins and topoisomerase IIα form central axes in the chromatid, while the rest are concentrated along it, around which the nucleosome fibers are wrapped.[46] Condensins use adenosine triphosphate (ATP-energy)-dependent processes that produce loops of DNA, a process known as loop extrusion (Figure 5). Condensin II makes larger loops, which end up forming a spiral staircase arrangement with a protruding loop at each step. Then, condensin I makes nested smaller loops within the condensin II loops.[66] To soundly sort out the topology of this, special proteins called topoisomerase IIα are recruited.[57,67]

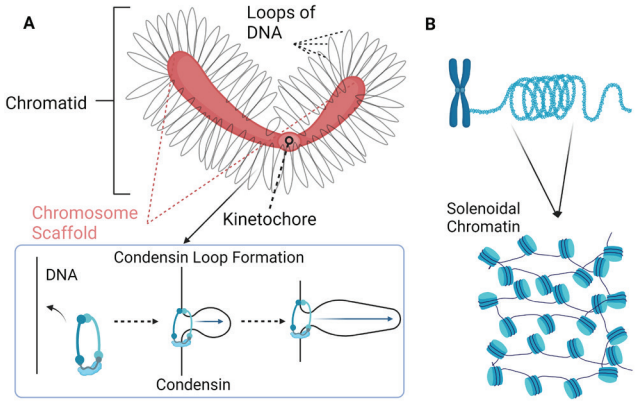


Figure 5: The Chromosome Scaffold and Condensin Loop Formation. **(A)** Schematic diagram of a chromatid with its loops of DNA visible, as well as the steps for condensin loop formation. **(B)** Shows the hierarchical folding model, particularly the solenoidal model of chromatin packaging. The radial loop model's dynamic nature can be seen in (A), while the solenoidal model in (B) is more restricted.

DNA topoisomerase IIα, also known as scaffolding protein 1 (Sc1), is localized at the axes of chromatids,[68] and although the exact role is unknown, it is assumed that Sc1 is essential in preventing the formation of knots and twists as the DNA gets shaped into the rod like arrays of loops representing mitotic chromosomes.[57,67] Sc1 does this by first cutting both strands of the target DNA, then passing an intact double helix through the break, and finally reconnecting the cut DNA. As the DNA is restored, the only difference between its initial and final state is the spatial orientation of the DNA.[69] Their depletion leads to long and thin chromosomes, similarly to condensin II depletion.[57] It is usually active not at prophase, but during late prophase/ the transition phase from prophase to metaphase, known as prometaphase.[70]

The other proteins at the central axes, condensins I and II, are ring-shaped proteins that have distinct yet overlapping functions, but condensin II is inside the nucleus during interphase, while condensin I is in the cytoplasm.[54,71] However, both do end up localizing at the chromatid axes during mitosis, since they are necessary for proper chromosome localization and topoisomerase IIα function.[57,72] They are ATPases as they bind to and hydrolyze ATP in their chemical reactions. The binding and hydrolysis of it regulate the opening and closing of the ring.[73] They shape the mitotic chromosomes by forming chromatin loops via loop extrusion.[74] They do this by binding onto DNA and then, while one part of the ring anchors, the opposite end slides back across the DNA using the motor activity, pulling the DNA along with it, and extruding it as a loop through the ring.[75]

While they both share the same SMC2/4 (Structural Maintenance of Chromosomes) dimer and appear similar, Condensins I and II have notable differences. For example, they have markedly different subunits. They are also different in that condensin I, but not II, requires chromokinesin KIF4A to localize to the mitotic chromosome axis.[76] Condensin II, but not I, requires the enzyme phosphatase 2A to associate with mitotic chromosomes.[77] Condensing activity also requires phosphorylation by kinases like Aurora B to relocate to the condensing chromatin and function there.[71,78] Additionally, condensin II is one of the first to contribute to looping and compacting the DNA from interphase during prophase, although there is a possibility of cohesin being active from prophase as well.[71] At the late stages of prophase, condensin I is brought in from the cytoplasm to the chromosome.[79]

Another essential protein for mitotic condensation, Cohesin is a protein made out of the SMC1/3 dimer that primarily connects sister chromatids during DNA replication,[80] but can interact with chromosomes and form loops during the cell cycle to compact them.[81] Recent studies have also shown that it influences the process of chromosome formation. Cohesin also plays a role in organizing interphase nuclei and regulates patterns of gene expression.[57] The cohesin complex also works with CTCF to organize interphase chromosomes into TADs as well as to help control gene expression.[52] Cohesins are released the most during prophase to act on the chromosome until the end of metaphase.[82] Their malfunction can cause genetic conditions, collectively known as cohesinopathies, including Robert's syndrome and Cornelia De Lange syndrome.

Although equally important, the SMC5/6 complex is elaborate, and thus its function is elusive. However, it is currently thought to play a role in associating with unusual DNA structures such as catenanes - molecules composed of at least 2 cyclic chemicals that are not chemically linked but need chemical link breakage to separate – as well as compacting the surrounding DNA and resolving topological tangles.[83] It can also be recruited by SAGA histone acetyltransferase to certain genes.[84] It seems to function during the S phase of the cell cycle to help replicate DNA and separate repetitive DNA, as well as potentially having roles in DNA repair and recombination. Depletion of it prior to entering S phase causes dramatic chromosome segregation defects.[57,85]

Chromokinesins, on the contrary, are a group of proteins that can move, and KIF4A is one of them. It usually plays a key role in forming the mitotic spindle and intercellular bridges,[86] but it is also connected to cancer if depleted, as it is required for regulating DNA damage responses.[87,88] It can bind to DNA and condensin I. The latter requires KIF4A to localize to the chromosomal axis;[89,90] however, it can only bind to condensin I after it has been phosphorylated.[91] Phosphorylation is also required for lateral chromosome compaction for KIF4A.[91] It has been suggested that KIF4A has a role in forming/stabilizing DNA loops, but has not been verified at the time of writing.[57]

It also works together with condensin and topoisomerase IIα to shape mitotic chromosomes.[90] Depletion causes a decrease in condensin I levels, thus causing short and fat chromosomes to form with defective structures.[90]

All six of these proteins are essential for mitotic condensation to occur, and evidence shows that when they are depleted, chromosomes have defects that can be significantly detrimental to humans, making understanding them illustrative of how errors in cell division can cause diseases. For example, short and fat chromosomes due to condensin I depletion can be found in certain cancers and can cause disrupted brain development.[92]

Chromatin Remodeling:

DNA bound by nucleosomes is very hard to bind to other factors, because it has to compete with the histones at the core of the nucleosomes. However, this closed structure can be opened. This opening, as well as many other processes, is collectively known as chromatin remodeling. There is a special class of proteins that use ATP called chromatin-remodeling complexes that carry these processes out (referred to as remodelers). They make sure of the proper density of nucleosomes and cooperate with site-specific transcription factors (TF) and histone-modifying proteins to move and eject histones from the nucleosome (sometimes entire nucleosomes), to allow the binding of TFs to DNA. Additionally, they help create special regions in the chromosome where canonical histones are replaced by variants. Remodelers are so crucial for regulating almost all chromosomal processes that the lack of them usually leads to many diseases, like cancer.[93,94]

The remodelers can be generally categorized into 4 distinct subfamilies: enzymes: imitation switch (ISWI), chromodomain helicase DNA-binding (CHD), switch/sucrose non-fermentable (SWI/SNF), and INO80.[94,95] While they might seem very different, recent studies have pointed out an important unifying aspect: all of them have an ATP-dependent movement of DNA along the histone surface to break histone-DNA contacts so that DNA can be driven along the histone. This is simply customized by the differing subfamilies to result in their specific actions. There are some other shared properties, such as preferring to bind to nucleosomes rather than naked DNA and having a singular ATPase subunit with a domain that allows interaction with other chromatin proteins.[94]

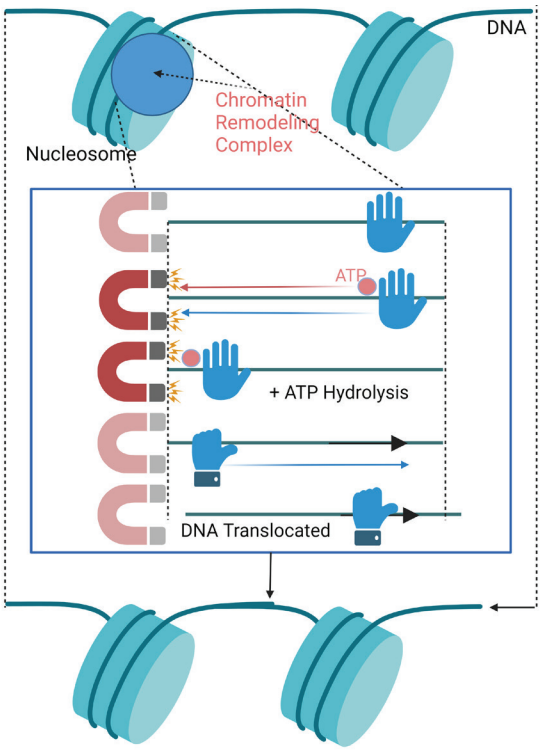


Figure 6: The mechanism of chromatin remodeling complex DNA translocation. Remodelers contain two lobes, where one is depicted as a U-magnet and the other as a hand in the diagram. Both lobes are first bound to the same strand of DNA, with one of them, implied as the magnet, ahead of the other. The hand is either closed or open, depending on how tightly the sequence of DNA that it is associated with binds to it. The lagging hand is moved towards the leading magnet when ATP is bound, but after its hydrolysis, the lagging hand moves backwards, pulling the bound and clamped DNA with it, moving 1-2 bp of DNA.[94] This mechanism is shared among the 4 categories of remodelers, although they all have distinct functions.

When at the nucleosome, and more specifically at the position SHL (Superhelical location) 2,[96] the remodeler binds near the entry site of DNA of the nucleosome and carries out its movement of DNA from the entry to the exit site (Figure 6).[97] Since the DNA is now more concentrated at the exit than the entry, the linker DNA is extended by 1-2 bp to resolve it. This results in overall movement of the histone octamer 1-2 bp along the DNA, which, through repetition, can be moved greater distances.[98] By modifying this process such that DNA is peeled off of adjacent nucleosomes, collision between adjacent ones, or causing major tension, a nucleosome may be ejected from its position. Furthermore, the strong tension can also break histone-DNA contacts to let nucleosome components be evicted and variants be added to restabilize.[94]

The different categories have their own purposes, such as when histone chaperones are carrying the histone polymers to the new DNA formed by DNA replication,[99] assembly remodelers like the ISWI and CHD are the ones to assist the histones and the histone octamer in forming a mature nucleosome. They then also regulate the distance between two nucleosomes so that it is fixed to form an array of nucleosomes. This also occurs at places where the nucleosomes have been ejected, like at sites of transcription. Since TFs need to bind to specific DNA regions called promoters that are blocked by

DOI: 10.36838/v7i10.114

the nucleosomes, the remodelers can move the necessary DNA region out of the blockage so it can be transcribed. While assembly remodelers indirectly silence genes, access remodelers like SWI/SNF are the ones mainly responsible for sliding histones, evicting nucleosome components like the H2A/H2B dimer, and evicting full nucleosomes for the purpose of making the chromatin more accessible to proteins and RNA, thus promoting gene expression.[100] Of note, they can also be used for gene repression. Additionally, they are also believed to be a central tumor suppressor.

INO80 remodelers, unlike the previous ones, are independent of replication and can remove a particular histone within a nucleosome and replace it with either a canonical or a variant histone. It is enriched at replication origins and DNA damage sites. Commonly, the histones H2A and H3 are replaced with the assistance of editing remodelers like SRCAP and p400, affecting factor recruitment, exclusion, and activity. INO80 can also relocate nucleosomes by as much as ~15 bp at a time by allowing the DNA loop to exit.[101]

All four of these remodelers' enzymatic activity can be regulated by either subunits that are next to the ATPase domain or by adjacent proteins via one of 'gating', ATP turnover, or 'coupling'. A substrate nucleosome can 'gate' the remodeler from working by altering DNA so that a change in shape is required for the remodeler to modify it. ATP turnover simply refers to the rate at which a remodeler uses ATP or goes through the cycle of binding and then hydrolyzing one ATP to the next. Coupling refers to either the probability that a usage of ATP results in movement of DNA, or the amount of DNA that is moved per one cycle of ATP hydrolysis.[94] Remodelers of all 4 subfamilies can also be affected and regulated by histone modifications and variations, as well as proteins like DNA-binding proteins.[102] They contain domains, such as bromodomains, bromo adjacent homology (BAH) domains, chromodomains, plant homeodomain (PHD) domains, Pro-Trp-Trp-Pro (PWWP) domains, and tryptophan-aspartic acid (WD40) domains. They either bind more or less to certain nucleosomes, depending on the variants, modifications, and nearby activators and repressors.

When their regulation meets errors, such as when SWI/SNF is altered to disrupt chromatin accessibility, oncogene activation and tumor suppressor silencing can occur, leading to cancer.[103] However, there are now clinical trials for inhibitors targeting remodelers to act as therapy against it.

Chromatin Organization and Gene Expression:

Chromatin is mainly separated into two compartments (3D regions of the nucleus), A and B. The A compartment is filled with transcriptionally active components like proteins and histone modifications; it is also associated with early replication. The B compartment is transcriptionally repressed and associated with late replication and includes inactivated genes and silencing modifications.[104,105] Additionally, there are 4 models proposed on how the larger-scale chromatin structure can affect transcription. This is because TF binding is a critical step in gene expression, and thus, the DNA being accessible is a significant part of gene regulation.

Firstly, steric occlusion is just the nucleosome or a series of nucleosome contacts blocking access to essential binding sites of TFs. This exclusion makes the concentration of TFs needed to significantly bind larger. In the extreme case that all binding sites are occupied, the concentration in that area is very low. If chromatin is bound to a gel, then regardless of the chemical properties, proteins are either small enough to enter and access the chromatin through the pores of the gel or too large. If the chromatin also forms a gel, it usually has a very well-defined pore size, leading to the blocking of entry of molecules depending on size. For example, it may shut out RNA Polymerase type II (RNAPII) due to its large size. Similarly, impenetrable crowding agents, such as chromatin, at a particular region can occupy so much solvent volume that it can decrease the concentration of soluble proteins by reducing the volume available for them.[8]

Next, Liquid-Liquid Phase Separation (LLPS) is a phenomenon where groups of molecules with weak interactions with each other separate into two liquid 'phases' with varying concentrations of major molecules to be more energy efficient for the cell than being in one mixture. It has been found that LLPS can occur for chromatin-associated proteins, including HP1α, linker histone, RNA-protein complexes, a diversity of other factors, and for chromatin itself. Therefore, condensed chromatin can exist in both a solid and a liquid state.[105] LLPS is also useful for partitioning as well as regulating transcription and accessibility, as it can prevent proteins and TFs from entering compartments, depending on its chemical properties, like charge. Additionally, the LLPS of chromatin compared to a bulk solution of it is compacted by 10,000-fold, showing that it can act to compact long-range chromatin, so much so that it is suggested that this state of chromatin is the natural 'ground state', although the specific structure at this state lacks scientific consensus.[8]

Examining further, we can see that there are many TFs with reduced residency times on the chromosome. For example, it has been seen that TFs like Sox2 can still access condensed mitotic chromosomes, albeit with decreased binding times due to the loss of stabilization by universal transcriptional inactivation.[106] The looping of chromatin by condensing, forming a spiral staircase, surprisingly does not significantly affect gene expression, and the local changes are most likely due to local factors like the availability of TFs and polymerases to bind DNA. There are some proposals on how condensin could mediate global silencing involving restricting promoter-enhancer interactions, but they will not be discussed further.[107] Pioneer TFs like FOXA1 can bind preferentially to nucleosomal DNA to initiate transcription and/or promote accessibility, while repressive TFs like CTCF do not.[108] All things considered, while the chromatin being condensed into chromosomes globally inhibits transcription, since TFs can still access and bind to chromatin, the genomic structure is more reflective of the local factors and regulatory elements rather than something universal.[8] Additionally, TFs can recruit remodelers and modifiers of histones to the nucleosome that they are near.[109]

Moreover, transcription and gene expression are processes that require initiation, elongation, and termination, and

so it has been observed to localize at less compact regions of chromatin. When genes on the massively silenced and chromatin/nucleosome-dense chromosome X avoid silencing, their transcription tends to localize away from the nearby inactive region of the chromosome. This has been shown to be true for the whole genome as RNAPII - the essential component of transcription that transcribes the DNA into mRNA - prefers to localize at compartments that are less densely packed with chromatin. Furthermore, histone depletion can also preferentially silence regions closer to large repressive domains like telomeres.[8] Interestingly, it has been shown through transcription inactivation studies that it is active transcription that regulates accessibility and short-range chromatin compaction instead of architectural scaffolding.[110,111] Also, transcription can increase accessibility of chromatin by destabilizing nucleosomes through the torsion created by the RNAPII moving along the DNA and processing it into mRNA, although this could be either reserved for a certain subset of genes or accumulated over long-term transcriptional activity.[8]

TADs also play a role in regulating gene expression. For example, the HoxD gene cluster - a cluster that is involved in many stages of vertebrate limb development - has two sequential bursts of gene activation, one for the preceding and central genes, then the next for the succeeding genes. This regulation between one to the next involves a switch in contacts between promoters and regulators, which define two distinct TADs: telomere-proximal TAD and centromere-proximal TAD (referred to as T-TAD and C-TAD). These TADs face each other with the HoxD gene cluster in between. Additionally, the chromatin structure defined by TADs was found likely to influence the phenotypes of certain diseases, such as the formation of limb anomalies.[112,113] Additionally, the chromatin organization at the stages of TADs can also regulate genes such as the sonic hedgehog gene (Shh), as the deletion of CTCF activity causes a 50% decrease in transcription of Shh. This is due to CTCF playing an important role in maintaining TADs. Thus, inhibiting it eliminates strong interactions between the promoter and enhancer.[114]

The ways that chromatin affects gene expression have significant implications for diseases. For instance, mutations in LLPS components can disrupt TAD boundaries, DNA repair, increase genomic instability, and promote oncogenic gene expression, thereby promoting cancer.[115]

■ Conclusion

Through discoveries from several decades of research in cell and molecular biology, scientists have been able to uncover a significant portion of information regarding chromatin and its form. As cell biological technologies improve, we hope that more can be uncovered in the spatial organization of chromatin in the nucleus and in how it affects biological processes.

This review has discussed such findings made by researchers on the topic of chromatin in the context of its organization. Specifically, it has covered the chemical composition of DNA and how the bond angles connecting the nucleic acid to the sugar phosphate backbone form minor and major grooves that proteins use to bind to precise locations. It has also reviewed the categories of chromatin from hetero to euchromatin and how it originates from Emil Heitz. Combining this with the histone octamer forms the nucleosome, the unit of chromatin. Additionally, I have outlined the process of octamer formation and how the chromatosome is formed. Although in the past, chromatin was thought to be hierarchically folded 30 nm fibers of DNA, the current consensus is that nucleosomes are arranged into irregular 10nm fibers. The dynamic irregular fibers are then looped around a chromosome scaffold that gives it its unique "X" shape during metaphase, composed of 6 essential proteins. The fibers of chromatin can be modified to allow for regularity or accessibility by inducing nucleosome maturation, translocation, eviction, or replacement. On the other hand, the decondensation of chromosomes during mitosis requires a combination of histone modifications. Finally, chromatin accessibility is regulated through steric occlusion, nuclear compartments, gel formation, and phase separation, as well as through the spatial organization of TADs and LADs.

In retrospect, ever since Emil Heitz's cytological staining to discover heterochromatin, innovations have allowed biologists to identify how *in vitro* artificial chromatin is organized into 30 nm fibers. Recent advances in assays and experimental techniques like cryo-electron microscopy, tomography, and partial decondensation have allowed us to identify the *in vivo* form of chromatin and how its "irregular" structure enables greater dynamism to control various genomic processes. Thus, in the following years, it is the hope that the remaining unanswered questions may be answered, such as how the key chromatin compaction proteins cooperate to produce the scaffolds to shape the chromosomes, the individual roles played by the different variants of proteins like chromokinesins, as well as how the precise mechanisms of chromatin remodeling complexes work to evict nucleosomes or their components.

To conclude, the study of DNA and how it is packaged remains essential to know due to how it affects gene expression and thus the phenotypes expressed. Yet, since the unit of genetic material, chromatin and its constituents, has been mainly uncovered, I believe there should be increased research attention on how these units contribute towards the larger processes in mitosis during interphase, as well as how they may be different in meiotic and polytene chromosomes. Polytene chromosomes, large chromosomes that occur from repeated rounds of DNA replication without daughter chromatid separation, are present in certain disorders, such as muscular dystrophy,[116] spontaneous abortions,[117] and, notably so in various tumor types in humans,[118,119] making their currently relatively unknown regulation and chromatin organization exceedingly important to research. Additionally, we have observed that the way chromatin impacts gene expression is interestingly usually dependent on its local conditions and, in certain cases, the permeability of the nuclear compartment. Thus, in the future, more on how chromatin organization through chromatin domains like TADs impacts specific gene loci and their expression can be potentially uncovered. However, there are still regions that we know little about, such as how we have only recently discovered the precise dynamics of the linker histone. Thus, there is

DOI: 10.36838/v7i10.114

information yet to be uncovered, such as how H1 recognizes the nucleosome and the details regarding the specific sites of interactions between the N and C-terminal domains of H1 within the chromatosome. Additionally, assays for identifying chromatin architecture at the short-range scale of a few nucleosomes have large potential for further development. Thus, as these tools are upgraded, our knowledge of the packaging of chromatin and how it relates to protein binding and gene expression can be elevated further.

By better understanding the form of DNA not only in mitotic but also meiotic cells and learning the interactions between the various proteins like condensin and cohesin, the probability of finding new measures to prevent certain diseases in the near future can be increased. Namely, diseases, such as cohesinopathies, including Roberts and Cornelia de Lange Syndrome, cancer, and developmental disorders like Down syndrome. Breast cancer has found chromatin accessibility among others as a new potential target for treatment.[120] Moreover, epigenetic drugs have recently demonstrated clinical success and have been approved for use against certain cancers.[121]

Fundamentally, according to the currently available data, chromatin is made up of grooved DNA spooled around a histone octamer and decorated by associated proteins and RNAs, which are packaged as irregular 10 nm fibers that are then looped around a protein scaffold to form the X-shaped chromosome. The nucleosome is then made of a histone octamer made of two of each of the core histones H2A, H2B, H3, and H4, that work together to bend the DNA around itself such that DNA can be compacted further. Additionally, from what has been demonstrated about chromatin and how it can affect a multitude of biological processes in the nucleus, there is very high potential in researching the local effects of chromatin on surrounding genes in narrower ranges. This is due to how essential transcription and gene expression are to the cell. However, as there is not as much research in the field done regarding it, I would strongly encourage studies in the area. This field of research has a very exciting future with more discoveries to come.

Acknowledgments

Special thanks to Kif Liakath-Ali for his mentoring of my project. I thank my family for their support in writing this paper. I thank Indigo Research for allowing me to write this review article. I thank the anonymous reviewers for their comments and feedback that helped to improve this paper.

All figures in this manuscript were made using Biorender.

References

1. Flemming W. *Zellsubstanz, Kern und Zelltheilung.* Vogel; 1882.
2. Paliulis L, Fabig G, Müller-Reichert T. The X chromosome still has a lot to reveal - revisiting Hermann Henking's work on firebugs. *J Cell Sci*. 2023;136(4):jcs260998. doi:10.1242/jcs.260998
3. Olins AL, Olins DE. Spheroid chromatin units (v bodies). *Science*. 1974;183(4122):330-332. doi:10.1126/science.183.4122.330
4. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin*. 2024;74(1):12-49. doi:10.3322/caac.21820
5. Zheng R, Wang S, Zhang S, et al. Global, regional, and national lifetime probabilities of developing cancer in 2020. *Sci Bull.* 2023;68(21):2620-2628. doi:10.1016/j.scib.2023.09.041
6. Levitt M. How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. *Proc Natl Acad Sci U S A*. 1978;75(2):640-644. doi:10.1073/pnas.75.2.640
7. Kowiel M, Brzezinski D, Gilski M, Jaskolski M. Conformation-dependent restraints for polynucleotides: the sugar moiety. *Nucleic Acids Res*. 2020;48(2):962-973. doi:10.1093/nar/gkz1122
8. Mansisidor AR, Risca VI. Chromatin accessibility: methods, mechanisms, and biological insights. *Nucleus*. 2022;13(1):236-276. doi:10.1080/19491034.2022.2143106
9. Ou HD, Phan S, Deerinck TJ, Thor A, Ellisman MH, O'Shea CC. ChromEMT: Visualizing 3D chromatin structure and compaction n interphase and mitotic cells. *Science*. 2017;357(6349):eaag0025. doi:10.1126/science.aag0025
10. Imai R, Nozaki T, Tani T, *et al*. Density imaging of heterochromatin in live cells using orientation-independent-DIC microscopy. *Mol Biol Cell.* 2017;28(23):3349-3359. doi:10.1091/mbc.E17-06-0359
11. Knight SC, Xie L, Deng W, *et al*. Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science*. 2015;350(6262):823-826. doi:10.1126/science.aac6572
12. Berger F. Emil Heitz, a true epigenetics pioneer. *Nat Rev Mol Cell Biol*. 2019;20(10):572. doi:10.1038/s41580-019-0161-z
13. Ambia-Garrido J, Vainrub A, Pettitt BM. A model for Structure and Thermodynamics of ssDNA and dsDNA Near a Surface: a Coarse Grained Approach. *Comput Phys Commun.* 2010;181(12):2001-2007. doi:10.1016/j.cpc.2010.08.029
14. Van Holde KE. *Chromatin. Springer*; 1989. doi:10.1007/978-1-4612-3490-6
15. Cutter AR, Hayes JJ. A brief review of nucleosome structure. *FEBS Lett.* 2015;589(20PartA):2914-2922. doi:10.1016/j.febslet.2015.05.016
16. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*. 1997;389(6648):251-260. doi:10.1038/38444
17. Wilkins BJ, Rall NA, Ostwal Y, et al. A cascade of histone modifications induces chromatin condensation in mitosis. *Science*. 2014;343(6166):77-80. doi:10.1126/science.1244508
18. Zhou K, Gaullier G, Luger K. Nucleosome structure and dynamics are coming of age. *Nat Struct Mol Biol*. 2019;26(1):3-13. doi:10.1038/s41594-018-0166-x
19. Beel AJ, Azubel M, Matteï PJ, Kornberg RD. Structure of Mitotic Chromosomes. *Mol Cell.* 2021;81(21):4369-4376.e3. doi:10.1016/j.molcel.2021.08.020
20. De Ambrosis A, Ferrari N, Bonassi S, Vidali G. Nucleosomal repeat length in active and inactive genes. *FEBS Lett*. 1987;225(1-2):120-122. doi:10.1016/0014-5793(87)81142-0
21. Tolstorukov MY, Colasanti AV, McCandlish DM, Olson WK, Zhurkin VB. A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. *J Mol Biol.* 2007;371(3):725-738. doi:10.1016/j.jmb.2007.05.048
22. Thåström A, Lowary PT, Widlund HR, Cao H, Kubista M, Widom J. Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J Mol Biol.* 1999;288(2):213-229. doi:10.1006/jmbi.1999.2686
23. Segal E, Fondufe-Mittendorf Y, Chen L, et al. A genomic code for nucleosome positioning. *Nature.* 2006;442(7104):772-778. doi:10.1038/nature04979
24. Struhl K, Segal E. Determinants of nucleosome positioning. *Nat Struct Mol Biol.* 2013;20(3):267-273. doi:10.1038/nsmb.2506

25. Raveh-Sadka T, Levo M, Shabi U, *et al.* Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nat Genet.* 2012;44(7):743-750. doi:10.1038/ng.2305
26. Jd A, J W. Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *J Mol Biol.* 2000;296(4). doi:10.1006/jmbi.2000.3531
27. Simpson RT. Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry.* 1978;17(25):5524-5531. doi:10.1021/bi00618a030
28. Thiriet C, Hayes JJ. Linker histone phosphorylation regulates global timing of replication origin firing. *J Biol Chem*. 2009;284(5):2823-2829. doi:10.1074/jbc.M805617200
29. Lu X, Wontakal SN, Kavi H, et al. Drosophila H1 regulates the genetic activity of heterochromatin by recruitment of Su(var)3-9. *Science*. 2013;340(6128):78-81. doi:10.1126/science.1234654
30. Shen X, Gorovsky MA. Linker histone H1 regulates specific gene expression but not global transcription *in vivo. Cell.* 1996;86(3):475-483. doi:10.1016/s0092-8674(00)80120-8
31. Zhang Y, Sun Z, Jia J, *et al*. Overview of Histone Modification. *Adv Exp Med Biol.* 2021;1283:1-16. doi:10.1007/978-981-15-8104-5_1
32. Greer EL, Shi Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet.* 2012;13(5):343-357. doi:10.1038/nrg3173
33. Ding L, Cao J, Lin W, *et al*. The Roles of Cyclin-Dependent Kinases in Cell-Cycle Progression and Therapeutic Strategies in Human Breast Cancer. *Int J Mol Sci.* 2020;21(6):1960. doi:10.3390/ijms21061960
34. Niu Y, Xu J, Sun T. Cyclin-Dependent Kinases 4/6 Inhibitors in Breast Cancer: Current Status, Resistance, and Combination Strategies. *J Cancer*. 2019;10(22):5504-5517. doi:10.7150/jca.32628
35. Cao L, Chen F, Yang X, Xu W, Xie J, Yu L. Phylogenetic analysis of CDK and cyclin proteins in premetazoan lineages. *BMC Evol Biol.* 2014;14:10. doi:10.1186/1471-2148-14-10
36. Robinson PJJ, An W, Routh A, *et al.* 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. *J Mol Biol.* 2008;381(4):816-825. doi:10.1016/j.jmb.2008.04.050
37. Ramadan K, Bruderer R, Spiga FM, *et al.* Cdc48/p97 promotes reformation of the nucleus by extracting the kinase Aurora B from chromatin. *Nature*. 2007;450(7173):1258-1262. doi:10.1038/nature06388
38. Antonin W, Neumann H. Chromosome condensation and decondensation during mitosis. *Curr Opin Cell Biol*. 2016;40:15-22. doi:10.1016/j.ceb.2016.01.013
39. Thadani R, Uhlmann F, Heeger S. Condensin, chromatin crossbarring and chromosome condensation. *Curr Biol CB.* 2012;22(23):R1012-1021. doi:10.1016/j.cub.2012.10.023
40. Schmitz MHA, Held M, Janssens V, *et al*. Live-cell imaging RNAi screen identifies PP2A–B55α and importin-β1 as key mitotic exit regulators in human cells. *Nat Cell Biol*. 2010;12(9):886-893. doi:10.1038/ncb2092
41. Luger K, Dechassa ML, Tremethick DJ. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol*. 2012;13(7):436-447. doi:10.1038/nrm3382
42. Finch JT, Klug A. Solenoidal model for superstructure in chromatin. *Proc Natl Acad Sci U S A*. 1976;73(6):1897-1901.
43. Fussner E, Strauss M, Djuric U, et al. Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. *EMBO Rep*. 2012;13(11):992. doi:10.1038/embor.2012.139
44. Eltsov M, MacLellan KM, Maeshima K, Frangakis AS, Dubochet J. Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. *Proc Natl Acad Sci U S A*. 2008;105(50):19732-19737. doi:10.1073/pnas.0810057105
45. Ma R, C M, Mf GP, M L, Mp C. Chromatin fibers are formed by heterogeneous groups of nucleosomes *in vivo. Cell*. 2015;160(6). doi:10.1016/j.cell.2015.01.054
46. Maeshima K, Imai R, Tamura S, Nozaki T. Chromatin as dynamic 10-nm fibers. *Chromosoma*. 2014;123(3):225-237. doi:10.1007/s00412-014-0460-2
47. Nishino Y, Eltsov M, Joti Y, et al. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *EMBO J*. 2012;31(7):1644-1653. doi:10.1038/emboj.2012.35
48. Maeshima K, Hihara S, Eltsov M. Chromatin structure: does the 30-nm fibre exist *in vivo*? *Curr Opin Cell Biol*. 2010;22(3):291-297. doi:10.1016/j.ceb.2010.03.001
49. Joti Y, Hikima T, Nishino Y, *et al*. Chromosomes without a 30-nm chromatin fiber. *Nucleus*. 2012;3(5):404-410. doi:10.4161/nucl.21222
50. Hirano T. At the heart of the chromosome: SMC proteins in action. *Nat Rev Mol Cell Biol*. 2006;7(5):311-322. doi:10.1038/nrm1909
51. K N, Ch H. The structure and function of SMC and kleisin complexes. *Annu Rev Biochem*. 2005;74. doi:10.1146/annurev.biochem.74.082803.133219
52. Nora EP, Lajoie BR, Schulz EG, et al. Spatial partitioning of the regulatory landscape of the X-inactivation center. *Nature.* 2012;485(7398):381-385. doi:10.1038/nature11049
53. Asakura S. On Interaction between Two Bodies Immersed in a Solution of Macromolecules - Astrophysics Data System. July 1954. Accessed October 15, 2024. https://ui.adsabs.harvard.edu/abs/1954JChPh..22.1255A/abstract
54. Ono T, Yamashita D, Hirano T. Condensin II initiates sister chromatid resolution during S phase. *J Cell Biol.* 2013;200(4):429-441. doi:10.1083/jcb.201208008
55. Guelen L, Pagie L, Brasset E, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature.* 2008;453(7197):948-951. doi:10.1038/nature06947
56. Y G, Rd G, R M, *et al*. The nuclear lamina and its functions in the nucleus. *Int Rev Cytol*. 2003;226. doi:10.1016/s0074-7696(03)01001-5
57. Paulson JR, Hudson DF, Cisneros-Soberanis F, Earnshaw WC. Mitotic chromosomes. *Semin Cell Dev Biol*. 2021;117:7-29. doi:10.1016/j.semcdb.2021.03.014
58. Giménez-Abián JF, Clarke DJ, Devlin J, et al. Premitotic chromosome individualization in mammalian cells depends on topoisomerase II activity. *Chromosoma.* 2000;109(4):235-244. doi:10.1007/s004120000065
59. Vagnarelli P, Hudson DF, Ribeiro SA, *et al*. Condensin and Repo-Man/PP1 co-operate in the regulation of chromosome architecture during mitosis. *Nat Cell Biol*. 2006;8(10):1133. doi:10.1038/ncb1475
60. St-Pierre J, Douziech M, Bazile F, *et al.* Polo kinase regulates mitotic chromosome condensation by hyperactivation of condensin DNA supercoiling activity. *Mol Cell*. 2009;34(4):416-426. doi:10.1016/j.molcel.2009.04.013
61. William E. Architecture of metaphase chromosomes and chromosome scaffolds. *J Cell Biol.* 1983;96(1):84-93.
62. Paulson JR, Vagnarelli P. Chromosomes and Chromatin. In: *eLS*. John Wiley & Sons, Ltd; 2011. doi:10.1002/9780470015902.a0005766.pub2

DOI: 10.36838/v7i10.114

63. Llères D, James J, Swift S, Norman DG, Lamond AI. Quantitative analysis of chromatin compaction in living cells using FLIM–FRET. *J Cell Biol*. 2009;187(4):481-496. doi:10.1083/jcb.200907029
64. Ohta S, Bukowski-Wills JC, Sanchez-Pulido L, et al. The Protein Composition of Mitotic Chromosomes Determined Using Multiclassifier Combinatorial Proteomics. *Cell*. 2010;142(5):810-821. doi:10.1016/j.cell.2010.07.047
65. Shintomi K, Inoue F, Watanabe H, Ohsumi K, Ohsugi M, Hirano T. Mitotic chromosome assembly despite nucleosome depletion in Xenopus egg extracts. *Science*. 2017;356(6344):1284-1287. doi:10.1126/science.aam9702
66. Walther N, Hossain MJ, Politi AZ, *et al*. A quantitative map of human Condensins provides new insights into mitotic chromosome architecture. *J Cell Biol*. 2018;217(7):2309-2328. doi:10.1083/jcb.201801048
67. Naumova N, Imakaev M, Fudenberg G, *et al*. Organization of the mitotic chromosome. *Science*. 2013;342(6161):948-953. doi:10.1126/science.1236083
68. Boy de la Tour E, Laemmli UK. The metaphase scaffold is helically folded: sister chromatids have predominantly opposite helical handedness. *Cell*. 1988;55(6):937-944. doi:10.1016/0092-8674(88)90239-5
69. Burden DA, Osheroff N. Mechanism of action of eukaryotic topoisomerase II and drugs targeted to the enzyme. *Biochim Biophys Acta BBA - Gene Struct Expr*. 1998;1400(1):139-154. doi:10.1016/S0167-4781(98)00132-8
70. Nielsen CF, Zhang T, Barisic M, Kalitsis P, Hudson DF. Topoisomerase IIα is essential for maintenance of mitotic chromosome structure. *Proc Natl Acad Sci U S A*. 2020;117(22):12131-12142. doi:10.1073/pnas.2001760117
71. Ono T, Fang Y, Spector DL, Hirano T. Spatial and Temporal Regulation of Condensins I and II in Mitotic Chromosome Assembly in Human Cells. *Mol Biol Cell*. 2004;15(7):3296-3308. doi:10.1091/mbc.E04-03-0242
72. Coelho PA, Queiroz-Machado J, Sunkel CE. Condensin-dependent localisation of topoisomerase II to an axial chromosomal structure is required for sister chromatid resolution during mitosis. *J Cell Sci*. 2003;116(Pt 23):4763-4776. doi:10.1242/jcs.00799
73. Hopfner KP, Tainer JA. Rad50/SMC proteins and ABC transporters: unifying concepts from high-resolution structures. *Curr Opin Struct Biol*. 2003;13(2):249-255. doi:10.1016/s0959-440x(03)00037-x
74. Hudson DF, Vagnarelli P, Gassmann R, Earnshaw WC. Condensin is required for nonhistone protein assembly and structural integrity of vertebrate mitotic chromosomes. *Dev Cell*. 2003;5(2):323-336. doi:10.1016/s1534-5807(03)00199-0
75. Kschonsak M, Merkel F, Bisht S, et al. Structural Basis for a Safety-Belt Mechanism That Anchors Condensin to Chromosomes. *Cell*. 2017;171(3):588-600.e24. doi:10.1016/j.cell.2017.09.008
76. Takahashi M, Wakai T, Hirota T. Condensin I-mediated mitotic chromosome assembly requires association with chromokinesin KIF4A. *Genes Dev*. 2016;30(17):1931-1936. doi:10.1101/gad.282855.116
77. Takemoto A, Maeshima K, Ikehara T, *et al*. The chromosomal association of condensin II is regulated by a noncatalytic function of PP2A. *Nat Struct Mol Biol*. 2009;16(12):1302-1308. doi:10.1038/nsmb.1708
78. Takemoto A, Kimura K, Yokoyama S, Hanaoka F. Cell cycle-dependent phosphorylation, nuclear localization, and activation of human condensin. *J Biol Chem*. 2004;279(6):4551-4559. doi:10.1074/jbc.M310925200
79. Hirota T, Gerlich D, Koch B, Ellenberg J, Peters JM. Distinct functions of condensin I and II in mitotic chromosome assembly. *J Cell Sci*. 2004;117(Pt 26):6435-6445. doi:10.1242/jcs.01604
80. Jeppsson K, Kanno T, Shirahige K, Sjögren C. The maintenance of chromosome structure: positioning and functioning of SMC complexes. *Nat Rev Mol Cell Biol*. 2014;15(9):601-614. doi:10.1038/nrm3857
81. Uhlmann F, Nasmyth K. Cohesion between sister chromatids must be established during DNA replication. *Curr Biol CB*. 1998;8(20):1095-1101. doi:10.1016/s0960-9822(98)70463-4
82. Sumara I, Vorlaufer E, Gieffers C, Peters BH, Peters JM. Characterization of Vertebrate Cohesin Complexes and Their Regulation in Prophase. *J Cell Biol*. 2000;151(4):749-762.
83. Serrano D, Cordero G, Kawamura R, *et al*. The Smc5/6 Core Complex Is a Structure-Specific DNA Binding and Compacting Machine. *Mol Cell*. 2020;80(6):1025-1038.e5. doi:10.1016/j.molcel.2020.11.011
84. Mahrik L, Stefanovie B, Maresova A, *et al*. The SAGA histone acetyltransferase module targets SMC5/6 to specific genes. *Epigenetics Chromatin*. 2023;16:6. doi:10.1186/s13072-023-00480-z
85. Venegas AB, Natsume T, Kanemaki M, Hickson ID. Inducible Degradation of the Human SMC5/6 Complex Reveals an Essential Role Only during Interphase. *Cell Rep*. 2020;31(3):107533. doi:10.1016/j.celrep.2020.107533
86. Hannabuss J, Lera-Ramirez M, Cade NI, Fourniol FJ, Nédélec F, Surrey T. Self-Organization of Minimal Anaphase Spindle Midzone Bundles. *Curr Biol CB*. 2019;29(13):2120-2130.e7. doi:10.1016/j.cub.2019.05.049
87. Song X, Zhang T, Wang X, *et al*. Distinct Diagnostic and Prognostic Values of Kinesin Family Member Genes Expression in Patients with Breast Cancer. *Med Sci Monit Int Med J Exp Clin Res*. 2018;24:9442-9464. doi:10.12659/MSM.913401
88. Wan Q, Shen Y, Zhao H, *et al*. Impaired DNA double-strand breaks repair by kinesin family member 4A inhibition renders human H1299 non-small-cell lung cancer cells sensitive to cisplatin. *J Cell Physiol*. 2019;234(7):10360-10371. doi:10.1002/jcp.27703
89. Almeida AC, Maiato H. Chromokinesins. *Curr Biol CB*. 2018;28(19):R1131-R1135. doi:10.1016/j.cub.2018.07.017
90. Samejima K, Samejima I, Vagnarelli P, *et al*. Mitotic chromosomes are compacted laterally by KIF4 and condensin and axially by topoisomerase IIα. *J Cell Biol*. 2012;199(5):755-770. doi:10.1083/jcb.201202155
91. Takata H, Madung M, Katoh K, Fukui K. Cdk1-dependent phosphorylation of KIF4A at S1186 triggers lateral chromosome compaction during early mitosis. *PloS One*. 2018;13(12):e0209614. doi:10.1371/journal.pone.0209614
92. Piskadlo E, Tavares A, Oliveira RA. Metaphase chromosome structure is dynamically maintained by condensin I-directed DNA (de)catenation. Sherratt D, ed. *eLife*. 2017;6:e26120. doi:10.7554/eLife.26120
93. Clapier CR, Cairns BR. Chromatin Remodeling Complexes. In: Workman JL, Abmayr SM, eds. *Fundamentals of Chromatin*. Springer; 2014:69-146. doi:10.1007/978-1-4614-8624-4_3
94. Clapier CR, Iwasa J, Cairns BR, Peterson CL. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol*. 2017;18(7):407-422. doi:10.1038/nrm.2017.26
95. Flaus A, Martin DMA, Barton GJ, Owen-Hughes T. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res*. 2006;34(10):2887. doi:10.1093/nar/gkl295

96. Längst G, Becker PB. ISWI induces nucleosome sliding on nicked DNA. *Mol Cell*. 2001;8(5):1085-1092. doi:10.1016/s1097-2765(01)00397-5
97. Nguyen VQ, Ranjan A, Stengel F, *et al*. Molecular Architecture of the ATP-Dependent Chromatin-Remodeling Complex SWR1. *Cell*. 2013;154(6):1220. doi:10.1016/j.cell.2013.08.018
98. Deindl S, Hwang WL, Hota SK, *et al*. ISWI remodelers slide nucleosomes with coordinated multi-base-pair entry steps and single-base-pair exit steps. *Cell*. 2013;152(3):442. doi:10.1016/j.cell.2012.12.040
99. Gurard-Levin ZA, Quivy JP, Almouzni G. Histone chaperones: assisting histone traffic and nucleosome dynamics. *Annu Rev Biochem*. 2014;83:487-517. doi:10.1146/annurev-biochem-060713-035536
100. Boeger H, Griesenbeck J, Strattan JS, Kornberg RD. Removal of promoter nucleosomes by disassembly rather than sliding *in vivo*. *Mol Cell*. 2004;14(5):667-673. doi:10.1016/j.molcel.2004.05.013
101. Eustermann S, Schall K, Kostrewa D, *et al*. Structural basis for nucleosome remodeling by the INO80 complex. *Nature*. 2018;556(7701):386. doi:10.1038/s41586-018-0029-y
102. Suganuma T, Workman JL. Signals and combinatorial functions of histone modifications. *Annu Rev Biochem*. 2011;80:473-499. doi:10.1146/annurev-biochem-061809-175347
103. Hodges C, Kirkland JG, Crabtree GR. The Many Roles of BAF (mSWI/SNF) and PBAF Complexes in Cancer. *Cold Spring Harb Perspect Med*. 2016;6(8):a026930. doi:10.1101/cshperspect.a026930
104. Harris HL, Gu H, Olshansky M, *et al*. Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. *Nat Commun*. 2023;14(1):3303. doi:10.1038/s41467-023-38429-1
105. Vilarrasa-Blasi R, Soler-Vila P, Verdaguer-Dot N, *et al*. Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. *Nat Commun*. 2021;12(1):651. doi:10.1038/s41467-020-20849-y
106. Teves SS, An L, Hansen AS, Xie L, Darzacq X, Tjian R. A dynamic mode of mitotic bookmarking by transcription factors. *eLife*. 2016;5:e22280. doi:10.7554/eLife.22280
107. Festuccia N, Owens N, Papadopoulou T, *et al*. Transcription factor activity and nucleosome organization in mitosis. *Genome Res*. 2019;29(2):250-260. doi:10.1101/gr.243048.118
108. Grossman SR, Zhang X, Wang L, *et al*. Systematic dissection of genomic features determining transcription factor binding and enhancer function. *Proc Natl Acad Sci U S A*. 2017;114(7):E1291-E1300. doi:10.1073/pnas.1621150114
109. Shukla A, Bhargava P. Regulation of tRNA gene transcription by the chromatin structure and nucleosome dynamics. *Biochim Biophys Acta BBA - Gene Regul Mech*. 2018;1861(4):295-309. doi:10.1016/j.bbagrm.2017.11.008
110. Jiang Y, Huang J, Lun K, et al. Genome-wide analyses of chromatin interactions after the loss of Pol I, Pol II, and Pol III. *Genome Biol*. 2020;21(1):158. doi:10.1186/s13059-020-02067-3
111. Neguembor MV, Martin L, Castells-García Á, *et al*. Transcription-mediated supercoiling regulates genome folding and loop formation. *Mol Cell*. 2021;81(15):3065. doi:10.1016/j.molcel.2021.06.009
112. Andrey G, Montavon T, Mascrez B, et al. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science*. 2013;340(6137):1234167. doi:10.1126/science.1234167
113. Kaiser VB, Semple CA. When TADs go bad: chromatin structure and nuclear organisation in human disease. *F1000Research*. 2017;6:F1000 Faculty Rev. doi:10.12688/f1000research.10792.1
114. Oudelaar AM, Higgs DR. The relationship between genome structure and function. *Nat Rev Genet*. 2021;22(3):154-168. doi:10.1038/s41576-020-00303-x
115. Tong X, Tang R, Xu J, et al. Liquid–liquid phase separation in tumor biology. *Signal Transduct Target Ther*. 2022;7(1):1-22. doi:10.1038/s41392-022-01076-x
116. Schmidt WM, Uddin MH, Dysek S, *et al*. DNA Damage, Somatic Aneuploidy, and Malignant Sarcoma Susceptibility in Muscular Dystrophies. *PLoS Genet*. 2011;7(4):e1002042. doi:10.1371/journal.pgen.1002042
117. Therman E, Denniston C, Sarto GE. Mitotic chiasmata in human diplochromosomes. *Hum Genet*. 1978;45(2):131-135. doi:10.1007/BF00286956
118. Therman E, Sarto GE, Buchler DA. The structure and origin of giant nuclei in human cancer cells. *Cancer Genet Cytogenet*. 1983;9(1):9-18. doi:10.1016/0165-4608(83)90019-5
119. Biesele JJ, Poyner H. Polytene Chromosomes in Two Mammary Carcinomas of the Human Subject. *Cancer Res*. 1943;3(11):779-783.
120. Wang X. Editorial: Chromatin architecture in gene regulation and disease. *Front Genet*. 2023;14:1254865. doi:10.3389/fgene.2023.1254865
121. Suraweera A, O'Byrne KJ, Richard DJ. Epigenetic drugs in cancer therapy. *Cancer Metastasis Rev*. 2025;44(1):37. doi:10.1007/s10555-025-10253-7

Author

Ryota is a Grade 10 international student at Crimson Global Academy. He is passionate about molecular biology and wants to become a researcher. He wants to go to a university in the United States to pursue biology as a major.

DOI: 10.36838/v7i10.114

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Role Of MECP2 Deficiency In Autism Spectrum Disorder Symptoms: A Pathway Analysis

Purva Sareen

Global Indian International School, 27 Punggol Field Walk, 828649, Singapore; purvasareen101@gmail.com
Mentor: Dr. Hamidreza Shaye

ABSTRACT: Methyl-CpG-binding protein-2 (*MECP2*) is a critical gene involved in neural development. Disruptions to the functioning of this gene have been associated with the risk of neurodevelopmental disorders, including Rett syndrome and autism spectrum disorder (ASD), characterized by social and cognitive deficits. In this paper, we test the hypothesis that *MECP2* deficiency may influence gene expression and disrupt specific molecular pathways, leading to ASD-related phenotypes. Using RNA-seq data from male mice *MECP2* mutant versus wild-type excitatory neurons, differential gene expression was mapped on a volcano plot. Genes that were significantly upregulated or downregulated were subjected to a Gene Ontology (GO) analysis to identify over-represented molecular pathways. Based on the findings, it was concluded that a deficiency in *MECP2* led to the dysregulation of vital neurotransmitter signaling pathways, including those for serotonin, dopamine, metabotropic glutamate, and muscarinic acetylcholine receptors, which show a strong correlation to ASD-related phenotypes. By integrating RNA-sequencing data with pathway analysis, this study aims to provide insights into how the reduction in the *MECP2* expression could contribute to the etiology of ASD, paving the way for targeted therapeutic strategies to alleviate the disease's behavioral symptoms.

KEYWORDS: Biomedical and Health Sciences, Genetics and Molecular Biology of Disease, Autism Spectrum Disorder, *MECP2* deficiency, Gene Ontology Analysis.

Introduction

Approximately 1 in every 100 children is diagnosed with Autism Spectrum Disorder (ASD) around the world,[1] making it one of the most common neurodevelopmental disorders. ASD is characterized by a range of communication and behavioral traits, including repetitive movements, avoidance of eye contact, and anxiety.[2] Due to the "anti-social" nature of ASD, patients with this disorder often face challenges with social interactions and slower processing speed,[3,4] making it notably difficult for them to adapt to social and environmental changes. Despite widespread awareness about ASD, its etiology remains complex and poorly understood, involving a range of environmental, genetic, and epigenetic factors.[5] (Figure 1)

Clinical trials suggest that one of the most substantial genetic risk factors for the disease is the presence of siblings with autism, with the recurrence rate of autism being approximately 2% to 8%.[6] Apart from familial clustering, other genetic and epigenetic contributors include advanced parental age, maternal immune activation (MIA), and neuroinflammation.[7–10] Environmental risk factors, such as heavy metal toxicity,[11] parental drug abuse,[12] and air pollutants,[13] are also gaining recognition for their potential implications on autism in the offspring.

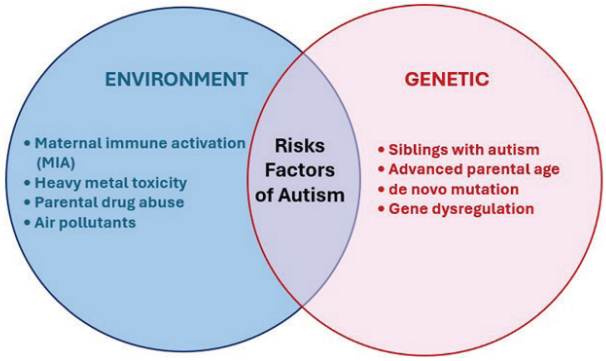


Figure 1: Venn Diagram of the genetic and environmental risk factors involved in the development of ASD. Image created using BioRender.com.

Due to the complexities of ASD, there is no curative or standard treatment plan for the disorder. The existing treatments for ASD are limited to early-year intervention with occupational, speech, and behavioral therapies to regulate the disorder's phenotypic characteristics, including aggression, self-injurious behavior (SIB), and severe tantrums.[14,15] However, medications can also be prescribed to children with comorbid challenging behaviors to reduce the intensity of these symptoms and improve social functioning.[16] To date, there are only two drugs that have been FDA-approved for the treatment of the behavioral symptoms of ASD in children: aripiprazole and risperidone.[17,18]

Ongoing research is spearheading advancements in targeted therapeutic strategies to expand the treatment landscape for ASD. One such drug that is gaining traction in the scientific community is buspirone (Buspar®),[19] an azapirone drug that

DOI: 10.36838/v7i10.127

acts as a partial agonist of the serotonin 5-HT_{1A} receptor. Buspirone works by binding to presynaptic 5-HT_{1A} receptors, a type of inhibitory autoreceptor, and partially activating them. (Figure 2) Over time, the drug causes increased serotonergic activity in the neurons by preventing the direct binding of serotonin and inducing desensitization of the somatodendritic autoreceptors.[20] Thus, the drug rebalances the levels of serotonin in the amygdala and other parts of the brain where 5-HT_{1A} receptors are prevalent, contributing to its anxiolytic effects. Buspirone is mainly used to treat generalized anxiety disorders (GADs) but also appears to be useful in other neurological disorders, including attention-deficit hyperactivity disorder and ASD.[21]

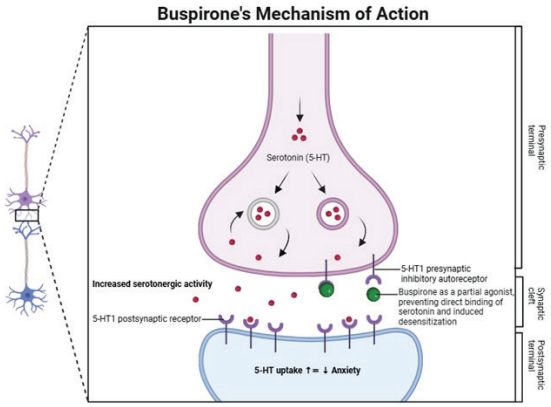


Figure 2: Mechanism of action of buspirone, a potent drug for anxiety in autism. Image created using BioRender.com.

Another novel therapeutic strategy includes positive allosteric modulators (PAMs) targeting the M4-muscarinic acetylcholine receptor (CHRM4). (Figure 3) The PAMs act as ligands that bind to an allosteric site of CHRM4, resulting in the potentiation of acetylcholine-induced responses and amplified downstream effects, including the inhibition of dopamine release.[22,23] A promising advantage of allosteric modulators is that the potentiation of the muscarinic acetylcholine receptors takes place only when and where the neurotransmitter acetylcholine is released, making sure to preserve the time and space pattern of acetylcholine action.[24,25] This drug mechanism is being explored for target therapeutic strategies in neurodevelopmental disorders like schizophrenia and ASD, where dysregulated dopamine signaling translates into typical phenotypes, including repetitive behavior and anxiety.[24,26,27]

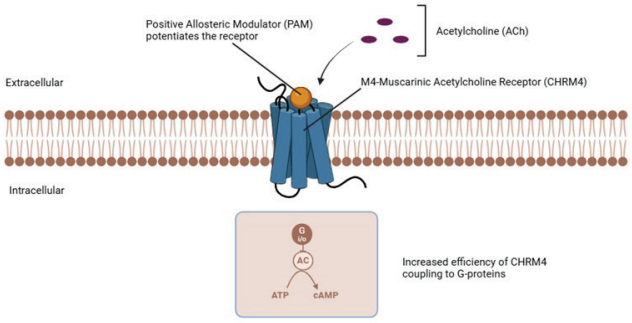


Figure 3: Mechanism of action of positive allosteric modulators (PAMs) for the M4-Muscarinic Acetylcholine Receptor (CHRM4). Image created using BioRender.com.

To strengthen evidence for these treatments, further molecular exploration of neurogenetic pathways is warranted. Among the numerous genes implicated in ASD pathway dysregulation, methyl-CpG-binding protein 2 (MECP2) emerges as a potential focus for understanding the molecular mechanisms underlying the symptoms of this disorder.[28,29] The *MECP2* gene, located on the X chromosome, is a critical epigenetic regulator that binds to methylated DNA and regulates gene expression and synaptic function, playing a central role in the postnatal development of the human brain.[30]

Recent studies have drawn correlations between mutations in *MECP2* and neurodevelopmental disorders, including Rett syndrome and broader ASD phenotypes such as social interaction deficits.[31] However, little research has been directed towards understanding the molecular correlations between *MECP2* gene deficiency and ASD symptoms. Therefore, this study seeks to address these gaps by investigating how *MECP2* deficiency can disrupt the functioning of specific molecular pathways, leading to the manifestation of ASD phenotypes. Furthermore, this study will explore how *MECP2* gene dysfunction alters potential molecular pathways relevant to ASD therapeutics and evaluate the impact of *MECP2* within the clinical landscape.

To fulfill the scope of this study, a secondary data analysis approach was employed, comprising of three main steps - 1) differential gene expression analysis of RNA-sequencing data from *MECP2*-deficient versus wild-type excitatory neurons in mice, 2) a subsequent Gene Ontology analysis for pathway enrichment, 3) exploration of how *MECP2* pathway dysregulation relates to therapeutic targets like buspirone and PAMs for CHRM4.

Methods

For this study, the RNA sequencing data were sourced from Supplementary Table 2 of an article published in Nature,[32] *"Characterization of human mosaic Rett syndrome brain tissue by single-nucleus RNA sequencing."* The respective dataset contained RNA-seq profiles from excitatory neurons of male mice *MECP2*–/y (1,230 cells) versus *MECP2*+/y (1,230 cells). Data normalization was not performed separately as the dataset was pre-processed and included identifiable columns for the mean expression in wild-type genes (meanWT) and knockout genes (meanKO), Log_2FC value, and p-value for each gene, allowing for minimal subsequent filtering. Next, the Benjamini-Hochberg procedure was used to calculate the false discovery rate (FDR) of the genes, which is a statistical measure that calculates the percentage likelihood of false positives in the gene set.[33] Genes with an FDR value of zero were marked as insignificant and removed from the data set using the 'DATA, FILTER' function on an Excel spreadsheet. This resulted in the reduction of the number of input genes from 25,284 to 8032. All the above-mentioned calculations and filtering were performed using Microsoft Excel.

Next, a volcano plot was created using the web-based tool Galaxy, usegalaxy.org,[34] to visualize the differential gene expression from the RNA-seq dataset. A volcano plot is a type of scatter graph that displays the magnitude of change in gene

expression due to a particular condition, which, in this study, is the absence of the *MECP2* gene in excitatory neurons.[35]

Thresholds were applied to the volcano plot to establish a focus on the respective data. A Log Fold Change threshold of $|Log_2FC| = 0.35$ was set, meaning genes with changes below this value were excluded from further analysis, denoted as the grey-dotted area of the graph. Log_2FC represents the magnitude and type of expression changes in a particular gene; a negative value displays down-regulation (cells are producing less of the coded protein), and a positive value displays up-regulation (cells are producing more of the coded protein). Additionally, a significance threshold of p-value < 0.0106 was applied to identify the genes with statistically robust differences. For further clarity, genes were colored if they passed these established thresholds – red for significantly upregulated genes and blue for significantly downregulated genes. Within this subset of colored genes, the top 15 genes with the smallest p-values were identified for further analysis. These genes were selected because they exhibited the most substantial changes in their expression due to MECP2 deficiency, serving as potent candidates for the following Gene Ontology (GO) analysis.[36]

Next, Gene Ontology (GO) analysis of the top 15 most significant genes was performed using the PANTHER Pathway toolkit available at geneontology.org. GO analysis is a method within bioinformatics that is used to identify cellular and molecular pathways that have been significantly over-represented within a selected set of genes. The criterion for pathway significance was 'adjusted p-value < 0.01'. The result of the GO analysis was then studied through literature reviews to draw connections between affected molecular pathways due to *MECP2* gene deficiency and behavioral symptoms of ASD.

Result and Discussion

The resultant volcano plot, generated using the web-based tool Galaxy, identified 547 differentially expressed genes (DEGs) between wildtype and *MECP2* knockdown mice from the total 8,032 genes, 254 downregulated and 293 upregulated. (Figure 4) With the established significance thresholds of $|Log_2FC| = 0.35$ and p-value < 0.0106 determined via FDR correction, 6.8% of the total gene collection was colored either red or blue, indicating increased or reduced gene expression, allowing for a focused analysis of only these genes that showed significant change in their expression. Among these, the top 15 genes with the most significant changes were identified and labelled on the graph: *Mecp2, Snap25, Junb, Arc, Nap1l5, Dynll1, Dynlrb1, Tsc22d1, Zc3hc1, Dynll2, Tspyl4, Camta1, Prkacb, Calm2, and Tmsb4x.* (Table 1)

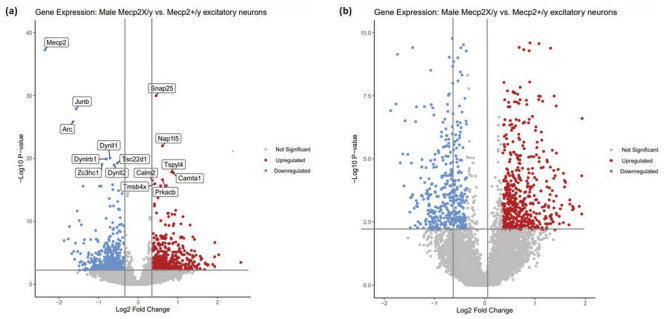


Figure 4: Volcano plot of differentially expressed genes between *MECP2*-deficient (MECP2–/y) and wild-type (MECP2+/y) excitatory neurons (1,230 cells per condition); (a) Full-scale volcano plot with Log_2FC on the x-axis and $-log_{10}$ P-value on the y-axis (0-40 range with top 15 most significantly differentially expressed genes (DEGs) labelled, (b) Volcano plot with y-axis range limited to 0-10 for clear visualization of DEG distribution and significance thresholds. Plots were generated using a web-based tool, Galaxy.

Table 1: Top 15 differentially expressed genes (DEGs) with the most significant upregulation or downregulation due to *MECP2* deficiency. The genes demonstrated the lowest p-values and highest fold changes, serving as primary candidates for subsequent gene ontology analysis. Table created using Microsoft Excel.

Sr No.	Gene Name	Log2FC (Fold Chain)	P-Value	Upregulated or Downregulated
1	Mecp2	-2.357881683	5.45E-38	Downregulated
2	Snap25	0.451994671	1.09E-30	Upregulated
3	Junb	-1.570122765	1.4E-28	Downregulated
4	Arc	-1.644369311	1.4E-26	Downregulated
5	Nap1l5	0.614110018	1.04E-22	Upregulated
6	Dynll1	-0.712450204	9.31E-21	Downregulated
7	Dynlrb1	-0.80382455	1.25E-20	Downregulated
8	Tsc22d1	-0.532035733	6.07E-20	Downregulated
9	Zc3hc1	-0.917429069	8.71E-20	Downregulated
10	Dynll2	-0.611337221	1.14E-19	Downregulated
11	Tspyl4	0.844798449	1.47E-18	Upregulated
12	Camta1	0.888635289	1.96E-18	Upregulated
13	Prkacb	0.614047402	2.44E-17	Upregulated
14	Calm2	0.354589978	3.18E-17	Upregulated
15	Tmsb4x	0.41942864	1.12E-16	Upregulated

Using the GO analysis and PANTHER toolkit, five pathways were identified as significantly over-represented by the top 15 genes. (Table 2)

Table 2: Molecular pathways were significantly enriched among the top 15 most significantly differentially expressed genes (DEGs). Pathways were identified using Gene Ontology enrichment analysis and the PANTHER pathway toolkit. All five identified pathways are significantly over-represented and are associated with key signaling systems disrupted by MECP2 deficiency. Table created using Microsoft Excel.

PANTHER Pathways	Set Size	Fold Enrichment	P-Value	Adjusted P-Value (FDR)
5HT1 type receptor mediated signaling pathway	2	60.66	4.89E-04	1.97E-02
Metabotropic glutamate receptor group II pathway	2	58.07	5.34E-04	1.72E-02
Dopamine receptor mediated signaling pathway	2	48.74	7.58E-04	2.03E-02
Muscarinic acetylcholine receptor 2 and 4 signaling pathway	2	45.49	8.69E-04	2.00E-02
Metabotropic glutamate receptor group III pathway	2	39.56	1.15E-03	2.31E-02

Notably, all five pathways showed a commonality in their involvement in G protein-coupled receptor (GPCR) signaling and neurotransmitter systems. This could indicate a potential shared mechanism of neurotransmitter dysregulation amongst the affected pathways, leading to the behavioral and neurological phenotypes of ASD patients. (Table 3) 5HT1 (serotonin), dopamine, metabotropic glutamate, and muscarinic acetylcholine receptors are GPCRs that trigger heterotrimeric G proteins to inhibit the adenylyl cyclase (AC) enzyme, which

DOI: 10.36838/v7i10.127

leads to reduced production of the second messenger molecule cyclic adenosine monophosphate (cAMP).[37] A reduction in these cAMP levels can lead to the dysfunction of protein kinase A (PKA) mediated biological processes in the brain, including disruptions in synaptic transmission and plasticity, neurogenesis, and apoptosis inhibition.[38] PKA is known for its role in phosphorylating multiple target proteins, including ion channels and transcription factors like cAMP response element-binding protein (CREB), which are crucial for long-term potentiation (LTP) and memory formation.[39,40] Ongoing research suggests a potential correlation between cAMP signaling dysregulation and ASD. For instance, studies have shown that cAMP production was found to be reduced in the platelets of patients with fragile X syndrome, a disease that often coexists with ASD.[41]

Table 3: Overview of the over-represented GPCR-mediated neurotransmitter pathways identified from the top DEGs. Dysregulation of these pathway mechanisms is implicated in ASD pathophysiology, drawing a potential correlation between MECP2 deficiency and the risk of ASD. Table created using Microsoft Excel.

Pathway	Receptor Family	Primary Molecule	Possible Contribution to ASD Phenotypes
5-HT1 Type Receptor Mediated Pathway	GPCR (5-HT1A, 5-HT1B, 5-HT1D, 5-HT1E, 5-HT1F)	Serotonin (5-HT)	Anxiety, impaired social behavior
Dopamine Receptor Mediated Pathway	GPCR (D1, D2, D3, D4, D5)	Dopamine	Social deficits, reduced reward processing, repetitive behaviors
Metabotropic Glutamate Receptor Pathway (Group II and III)	GPCR (mGluR2, mGluR3, mGluR4, mGluR6-8)	Glutamate	Learning deficits, anxiety
Muscarinic Acetylcholine Receptor Pathway (M2 and M4)	GPCR (M2 and M4)	Acetylcholine	Learning deficits, anxiety, weak memory capacity

Apart from the commonality of AC inhibition and cAMP reduction, these pathways also display other vital roles in the human nervous system. Dysregulation of the aforementioned pathways due to *MECP2* deficiency can result in the development of emotional and behavioral abnormalities often seen in ASD.

Serotonergic Pathway in ASD:

As indicated in Table 3, the GO enrichment analysis identified a Fold Enrichment value of 60.66 and a corrected p-value of 0.0172 for the 5HT1 type receptor-mediated signaling pathway, making it the most significantly overrepresented pathway from the input gene set. The 5-HT1 serotonin receptors are a group of receptors, mainly found in the brain, that bind to serotonin, a neurotransmitter that is known to regulate mood, memory, and learning.[42] During brain maturation at early stages of life, 5-HT1 plays a key role as a neurotransmitter and a growth factor that regulates neural development, synapse formation, and cortical arrangement.[43] A study done by the Center for Neurobiology and Behavior, Columbia University, concluded that 5-HT_{1A} receptor knockout mice exhibited increased anxiety-like behaviors in a series of behavioral tests, suggesting that 5-HT_{1A} receptors are involved in the regulation of exploratory and fear-related behaviors.[44] Thus, it can be deduced that disruptions in the 5-HT1 type receptor-mediated signaling pathway could lead to behavioral deficits such as anxiety and impaired social interactions, phenotypes often seen in ASD patients.

Dopaminergic Pathway in ASD:

As indicated in Table 3, the GO enrichment analysis identified a Fold Enrichment value of 48.74 and a corrected p-value of 0.0203 for the dopamine receptor-mediated signaling pathway. Dopamine is a vital neurotransmitter that regulates a list of physiological processes, including motivation, reward, and cognition.[45,46] Dopamine receptors have been recognized as key players in early brain development through their regulation of neuronal differentiation and axonal outgrowth. Recent studies conclude that dopamine receptor activation aids in the trafficking and functioning of AMPA and NMDA receptors, key mediators of excitatory neurotransmission.[47] The dysregulation of dopamine signaling has been implicated in various neurodevelopmental disorders (NDDs), including schizophrenia and autism spectrum disorder. One of the primary outcomes of dopamine receptor dysfunction is the disruption of synaptic plasticity, particularly within the striatum and prefrontal cortex – two regions of the brain that are enriched with dopamine receptors and are responsible for behavioral attributes, including social interaction.[48] Thus, disrupted dopaminergic pathways might contribute to the behavioral actualization of ASD, including low reward value from social stimuli, i.e., social anhedonia and disrupted sensorimotor processing.[49]

Overall, the paper's findings suggest that *MECP2* deficiency disrupts critical pathways that are responsible for neurodevelopment and synaptogenesis, including the regulation of GPCR signaling and cAMP levels. Thus, these molecular dysregulations – influenced by *MECP2* – contribute to ASD phenotypes, including anxiety and stereotypies. These insights align with the mechanism of action of emerging therapeutic strategies such as buspirone, which targets the 5-HT_{1A} serotonin receptors, and CHRM4 positive allosteric modulators, both of which aim to regulate the neurotransmitter systems that were disrupted by *MECP2* deficiency. This strengthens the rationale for further exploring such drugs in the treatment of ASD symptoms.

However, a limitation of this study is that conclusions on the affected molecular pathways are based on the computational analysis of Gene Ontology. The insights provided need to be validated by future experiments with *MECP2* gene knockout mice to assess the likelihood of these results. Moreover, it is noteworthy that it is highly unlikely that a single genetic insult is responsible for all cases of biologically complex and clinically heterogeneous neurodevelopmental disorders such as ASD.

■ Conclusion

In summary, the initial hypothesis of the study stated that reduced *MECP2* gene expression stands as a potential epigenetic factor for the development of ASD and its behavioral phenotypes. The pathway analyses applied in this study serve as a proof-of-concept for this hypothesis, exploring how reduced *MECP2* expression can disrupt the functionality of specific GPCRs and neurotransmitter signaling pathways, contributing to the phenotypic abnormalities of the disorder. Our results align with the mechanisms of novel ASD therapeutics, such as buspirone and CHRM4 positive allosteric modulators, and emphasize the importance of focusing drug targets for ASD in the serotonergic and dopaminergic pathways, which these drugs target. The findings of this study can be further developed upon via experimental validation in *MECP2*-deficient

animal models and large-scale transcriptomic analyses with larger gene sets.

Further studies on potential genetic and epigenetic risk factors, including *MECP2* gene deficiency, are crucial in expanding treatment methods for the disorder's behavioral and neurological phenotypes. Lack of etiological understanding of ASD has restricted treatment options for patients with this neurodevelopmental disorder. Moreover, advancing in target therapeutics through pre-clinical and clinical trials of drug efficacy, including buspirone and PAMs for CHRM4, must also be prioritized to cater to the growing prevalence of ASD worldwide.

Acknowledgments

I would like to thank my mentors, Dr. Hamidreza Shaye and Dr. Lauren Tetz from the IRIS Program, for their constant guidance throughout the research and publication process. I would also like to thank BioRender for providing an effective tool to illustrate specific mechanisms and pathways in my paper.

References

1. Zeidan, J.; Fombonne, E.; Scorah, J.; Ibrahim, A.; Durkin, M. S.; Saxena, S.; Yusuf, A.; Shih, A.; Elsabbagh, M. Global Prevalence of Autism: A Systematic Review Update. *Autism Research* **2022**, *15* (5), 778–790. https://doi.org/10.1002/aur.2696.
2. Hodges, H.; Fealko, C.; Soares, N. Autism Spectrum Disorder: Definition, Epidemiology, Causes, and Clinical Evaluation. *Transl Pediatr* **2020**, *9* (S1), S55–S65. https://doi.org/10.21037/tp.2019.09.09.
3. Garfin, D. G.; Lord, C. Communication as a Social Problem in Autism. In *Social Behavior in Autism*; Schopler, E., Mesibov, G. B., Eds.; Springer US: Boston, MA, 1986; pp 133–151. https://doi.org/10.1007/978-1-4899-2242-7_7.
4. Linnenbank, M.; Feldmann, R.; Schulte-Körne, G.; Beimdiek, S.; Strittmatter, E. Children with Autism Spectrum Disorder of All Ages, Levels of Symptom Severity and General Cognitive Ability Display Low Processing Speed Index Scores Warranting Special Educational Assistance. *J Autism Dev Disord* **2022**, *52* (8), 3668–3675. https://doi.org/10.1007/s10803-021-05249-5.
5. Oztenekecioglu, B.; Mavis, M.; Osum, M.; Kalkan, R. Genetic and Epigenetic Alterations in Autism Spectrum Disorder. *Glob Med Genet* **2021**, *08* (04), 144–148. https://doi.org/10.1055/s-0041-1735540.
6. Muhle, R.; Trentacoste, S. V.; Rapin, I. The Genetics of Autism. *Pediatrics* **2004**, *113* (5), e472-486. https://doi.org/10.1542/peds.113.5.e472.
7. Wu, S.; Wu, F.; Ding, Y.; Hou, J.; Bi, J.; Zhang, Z. Advanced Parental Age and Autism Risk in Children: A Systematic Review and Meta-Analysis. *Acta Psychiatrica Scandinavica* **2017**, *135* (1), 29–41. https://doi.org/10.1111/acps.12666.
8. Han, V. X.; Patel, S.; Jones, H. F.; Dale, R. C. Maternal Immune Activation and Neuroinflammation in Human Neurodevelopmental Disorders. *Nat Rev Neurol* **2021**, *17* (9), 564–579. https://doi.org/10.1038/s41582-021-00530-8.
9. Tartaglione, A. M.; Villani, A.; Ajmone-Cat, M. A.; Minghetti, L.; Ricceri, L.; Pazienza, V.; De Simone, R.; Calamandrei, G. Maternal Immune Activation Induces Autism-like Changes in Behavior, Neuroinflammatory Profile and Gut Microbiota in Mouse Offspring of Both Sexes. *Transl Psychiatry* **2022**, *12* (1), 384. https://doi.org/10.1038/s41398-022-02149-9.
10. Choi, G. B.; Yim, Y. S.; Wong, H.; Kim, S.; Kim, H.; Kim, S. V.; Hoeffer, C. A.; Littman, D. R.; Huh, J. R. The Maternal Interleukin-17a Pathway in Mice Promotes Autism-like Phenotypes in Offspring. *Science* **2016**, *351* (6276), 933–939. https://doi.org/10.1126/science.aad0314.
11. Błażewicz, A.; Grabrucker, A. M. Metal Profiles in Autism Spectrum Disorders: A Crosstalk between Toxic and Essential Metals. *IJMS* **2022**, *24* (1), 308. https://doi.org/10.3390/ijms24010308.
12. Butwicka, A.; Långström, N.; Larsson, H.; Lundström, S.; Serlachius, E.; Almqvist, C.; Frisén, L.; Lichtenstein, P. Increased Risk for Substance Use-Related Problems in Autism Spectrum Disorders: A Population-Based Cohort Study. *J Autism Dev Disord* **2017**, *47* (1), 80–89. https://doi.org/10.1007/s10803-016-2914-2.
13. Pagalan, L.; Bickford, C.; Weikum, W.; Lanphear, B.; Brauer, M.; Lanphear, N.; Hanley, G. E.; Oberlander, T. F.; Winters, M. Association of Prenatal Exposure to Air Pollution With Autism Spectrum Disorder. *JAMA Pediatr* **2019**, *173* (1), 86. https://doi.org/10.1001/jamapediatrics.2018.3101.
14. Erickson, C. A.; Posey, D. J.; Stigler, K. A.; McDougle, C. J. Pharmacologic Treatment of Autism and Related Disorders. *Pediatr Ann* **2007**, *36* (9), 575–585. https://doi.org/10.3928/0090-4481-20070901-09.
15. Zwaigenbaum, L.; Bauman, M. L.; Choueiri, R.; Kasari, C.; Carter, A.; Granpeesheh, D.; Mailloux, Z.; Smith Roley, S.; Wagner, S.; Fein, D.; Pierce, K.; Buie, T.; Davis, P. A.; Newschaffer, C.; Robins, D.; Wetherby, A.; Stone, W. L.; Yirmiya, N.; Estes, A.; Hansen, R. L.; McPartland, J. C.; Natowicz, M. R. Early Intervention for Children With Autism Spectrum Disorder Under 3 Years of Age: Recommendations for Practice and Research. *Pediatrics* **2015**, *136* Suppl 1 (Suppl 1), S60-81. https://doi.org/10.1542/peds.2014-3667E.
16. Alsayouf, H. A.; Talo, H.; Biddappa, M. L.; De Los Reyes, E. Risperidone or Aripiprazole Can Resolve Autism Core Signs and Symptoms in Young Children: Case Study. *Children* **2021**, *8* (5), 318. https://doi.org/10.3390/children8050318.
17. Mano-Sousa, B. J.; Pedrosa, A. M.; Alves, B. C.; Galduróz, J. C. F.; Belo, V. S.; Chaves, V. E.; Duarte-Almeida, J. M. Effects of Risperidone in Autistic Children and Young Adults: A Systematic Review and Meta-Analysis. *CN* **2021**, *19* (4), 538–552. https://doi.org/10.2174/1570159X18666200529151741.
18. Hutchinson, J.; Folawemi, O.; Bittla, P.; Kaur, S.; Sojitra, V.; Zahra, A.; Khan, S. The Effects of Risperidone on Cognition in People With Autism Spectrum Disorder: A Systematic Review. *Cureus* **2023**. https://doi.org/10.7759/cureus.45524.
19. Chessick, C. A.; Allen, M. H.; Thase, M. E.; Batista Miralha Da Cunha, A. A.; Kapczinski, F.; Silva De Lima, M.; Dos Santos Souza, J. J. Azapirones for Generalized Anxiety Disorder. *Cochrane Database of Systematic Reviews* **2006**, *2015* (6). https://doi.org/10.1002/14651858.CD006115.
20. *Buspirone*. https://go.drugbank.com/drugs/DB00490 (accessed 2025-02-06).
21. Wilson, T. K.; Tripp, J. Buspirone. In StatPearls; StatPearls Publishing: Treasure Island (FL), 2025.
22. *Positive Allosteric Modulator - an overview | ScienceDirect Topics*. https://www.sciencedirect.com/topics/pharmacology-toxicology-and-pharmaceutical-science/positive-allosteric-modulator (accessed 2025-01-23).
23. Vuckovic, Z.; Wang, J.; Pham, V.; Mobbs, J. I.; Belousoff, M. J.; Bhattarai, A.; Burger, W. A.; Thompson, G.; Yeasmin, M.; Nawaratne, V.; Leach, K.; Van Der Westhuizen, E. T.; Khajehali, E.; Liang, Y.-L.; Glukhova, A.; Wootten, D.; Lindsley, C. W.; Tobin, A.; Sexton, P.; Danev, R.; Valant, C.; Miao, Y.; Christopoulos, A.; Thal, D. M. Pharmacological Hallmarks of Allostery at the M4

DOI: 10.36838/v7i10.127

Muscarinic Receptor Elucidated through Structure and Dynamics. *eLife* **2023**, *12*, e83477. https://doi.org/10.7554/eLife.83477.
24. Kruse, A. C.; Kobilka, B. K.; Gautam, D.; Sexton, P. M.; Christopoulos, A.; Wess, J. Muscarinic Acetylcholine Receptors: Novel Opportunities for Drug Development. *Nat Rev Drug Discov* **2014**, *13* (7), 549–560. https://doi.org/10.1038/nrd4295.
25. Cao, A.-M.; Quast, R. B.; Fatemi, F.; Rondard, P.; Pin, J.-P.; Margeat, E. Allosteric Modulators Enhance Agonist Efficacy by Increasing the Residence Time of a GPCR in the Active State. *Nat Commun* **2021**, *12* (1), 5426. https://doi.org/10.1038/s41467-021-25620-5.
26. Jakubík, J. Allosteric Modulation of Muscarinic Acetylcholine Receptors.
27. Chambers, N. E.; Millett, M.; Moehle, M. S. The Muscarinic M4 Acetylcholine Receptor Exacerbates Symptoms of Movement Disorders. *Biochemical Society Transactions* **2023**, *51* (2), 691–702. https://doi.org/10.1042/BST20220525.
28. Nagarajan, R.; Hogart, A.; Gwye, Y.; Martin, M. R.; LaSalle, J. M. Reduced MeCP2 Expression Is Frequent in Autism Frontal Cortex and Correlates with Aberrant MECP2 Promoter Methylation. *Epigenetics* **2006**, *1* (4), 172–182. https://doi.org/10.4161/epi.1.4.3514.
29. Gonzales, M. L.; LaSalle, J. M. The Role of MeCP2 in Brain Development and Neurodevelopmental Disorders. *Curr Psychiatry Rep* **2010**, *12* (2), 127–134. https://doi.org/10.1007/s11920-010-0097-7.
30. Zimmermann, C.; Hoffmann, A.; Raabe, F.; Spengler, D. Role of Mecp2 in Experience-Dependent Epigenetic Programming. *Genes* **2015**, *6* (1), 60–86. https://doi.org/10.3390/genes6010060.
31. Goffin, D.; Allen, M.; Zhang, L.; Amorim, M.; Wang, I.-T. J.; Reyes, A.-R. S.; Mercado-Berton, A.; Ong, C.; Cohen, S.; Hu, L.; Blendy, J. A.; Carlson, G. C.; Siegel, S. J.; Greenberg, M. E.; Zhou, Z. Rett Syndrome Mutation MeCP2 T158A Disrupts DNA Binding, Protein Stability and ERP Responses. *Nat Neurosci* **2012**, *15* (2), 274–283. https://doi.org/10.1038/nn.2997.
32. Renthal, W.; Boxer, L. D.; Hrvatin, S.; Li, E.; Silberfeld, A.; Nagy, M. A.; Griffith, E. C.; Vierbuchen, T.; Greenberg, M. E. Characterization of Human Mosaic Rett Syndrome Brain Tissue by Single-Nucleus RNA Sequencing. *Nature Neuroscience* **2018**, *21* (12), 1670–1679. https://doi.org/10.1038/s41593-018-0270-6.
33. Chen, S.-Y.; Feng, Z.; Yi, X. A General Introduction to Adjustment for Multiple Comparisons. *J. Thorac. Dis.* **2017**, *9* (6), 1725–1729. https://doi.org/10.21037/jtd.2017.05.34.
34. *Galaxy*. https://usegalaxy.org/ (accessed 2025-01-26).
35. Li, W. Volcano Plots in Analyzing Differential Expressions with mRNA Microarrays. *Journal of bioinformatics and computational biology* **2012**, *10*, 1231003. https://doi.org/10.1142/S0219720012310038.
36. *GO citation policy and license*. Gene Ontology Resource. http://geneontology.org/docs/go-citation-policy/ (accessed 2025-01-26).
37. Kelley, D.; Bhattacharyya, A.; Lahvis, G.; Yin, J.; Malter, J.; Davidson, R. The Cyclic AMP Phenotype of Fragile X and Autism. *Neuroscience & Biobehavioral Reviews* **2008**, *32* (8), 1533–1543. https://doi.org/10.1016/j.neubiorev.2008.06.005.
38. Chen, D.; Wang, J.; Cao, J.; Zhu, G. cAMP-PKA Signaling Pathway and Anxiety: Where Do We Go Next? *Cellular Signalling* **2024**, *122*, 111311. https://doi.org/10.1016/j.cellsig.2024.111311.
39. Kida, S. A Functional Role for CREB as a Positive Regulator of Memory Formation and LTP. *Exp Neurobiol* **2012**, *21* (4), 136–140. https://doi.org/10.5607/en.2012.21.4.136.
40. Costa-Mattioli, M.; Sossin, W. S.; Klann, E.; Sonenberg, N. Translational Control of Long-Lasting Synaptic Plasticity and Memory. *Neuron* **2009**, *61* (1), 10–26. https://doi.org/10.1016/j.neuron.2008.10.055.
41. Berry-Kravis, E.; Hicar, M.; Ciurlionis, R. Reduced Cyclic AMP Production in Fragile X Syndrome: Cytogenetic and Molecular Correlations. *Pediatr Res* **1995**, *38* (5), 638–643. https://doi.org/10.1203/00006450-199511000-00002.
42. Kanova, M.; Kohout, P. Serotonin—Its Synthesis and Roles in the Healthy and the Critically Ill. *IJMS* **2021**, *22* (9), 4837. https://doi.org/10.3390/ijms22094837.
43. Bonnin, A.; Levitt, P. Fetal, Maternal, and Placental Sources of Serotonin and New Implications for Developmental Programming of the Brain. *Neuroscience* **2011**, *197*, 1–7. https://doi.org/10.1016/j.neuroscience.2011.10.005.
44. Ramboz, S.; Oosting, R.; Amara, D. A.; Kung, H. F.; Blier, P.; Mendelsohn, M.; Mann, J. J.; Brunner, D.; Hen, R. Serotonin Receptor 1A Knockout: An Animal Model of Anxiety-Related Disorder. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95* (24), 14476–14481. https://doi.org/10.1073/pnas.95.24.14476.
45. Speranza, L.; Di Porzio, U.; Viggiano, D.; De Donato, A.; Volpicelli, F. Dopamine: The Neuromodulator of Long-Term Synaptic Plasticity, Reward and Movement Control. *Cells* **2021**, *10* (4), 735. https://doi.org/10.3390/cells10040735.
46. Juárez Olguín, H.; Calderón Guzmán, D.; Hernández García, E.; Barragán Mejía, G. The Role of Dopamine and Its Dysfunction as a Consequence of Oxidative Stress. *Oxidative Medicine and Cellular Longevity* **2016**, *2016* (1), 9730467. https://doi.org/10.1155/2016/9730467.
47. Neve, K. A.; Seamans, J. K.; Trantham-Davidson, H. Dopamine Receptor Signaling. *Journal of Receptors and Signal Transduction* **2004**, *24* (3), 165–205. https://doi.org/10.1081/RRS-200029981.
48. Samaco, R. C.; Mandel-Brehm, C.; Chao, H.-T.; Ward, C. S.; Fyffe-Maricich, S. L.; Ren, J.; Hyland, K.; Thaller, C.; Maricich, S. M.; Humphreys, P.; Greer, J. J.; Percy, A.; Glaze, D. G.; Zoghbi, H. Y.; Neul, J. L. Loss of MeCP2 in Aminergic Neurons Causes Cell-Autonomous Defects in Neurotransmitter Synthesis and Specific Behavioral Abnormalities. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (51), 21966–21971. https://doi.org/10.1073/pnas.0912257106.
49. Blum, K.; Bowirrat, A.; Sunder, K.; Thanos, P. K.; Hanna, C.; Gold, M. S.; Dennen, C. A.; Elman, I.; Murphy, K. T.; Makale, M. T. Dopamine Dysregulation in Reward and Autism Spectrum Disorder. *Brain Sciences* **2024**, *14* (7), 733. https://doi.org/10.3390/brainsci14070733.

■ Author

Purva Sareen is currently an International Baccalaureate (IB) Year 1 student at Global Indian International School, Singapore. She is particularly interested in neurodevelopmental disorders and aspires to study medicine in the near future.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

The Effect of Universal School Lunches on Students' Academic Performance.

Sascha Pirghibi

Sage Hill High School, 20402 Newport Coast Dr, Newport Coast, CA 92657, USA; spirghibi@gmail.com

ABSTRACT: Can free school lunches accomplish more than merely providing a meal to students? This research project examines the idea that providing free meals at school could be an effective approach for improving academic achievement and leveling the playing field for all kids, regardless of their financial situation. This research answers whether receiving free lunch increases academic performance in students and decreases absenteeism rates in schools. While hypothesizing, it was proposed that if students are eating at least one fulfilling meal every day, their academic performance should slightly increase, and absenteeism should decrease. An analytical approach is going to be used to compare academic outcomes between schools that established free meal programs at different dates to test the hypothesis that free school lunches improve academic performance and reduce absenteeism. It is important to separate the effects of the lunch program from other variables by adjusting for attendance records and results from standardized tests. The academic performance results highlight that student performance is not affected by free school lunches, as there is no significant change before and after school lunches were freely provided. Absenteeism did rise after the change to free school lunches, but with the overall lack of change in student academic performance, absenteeism does not affect the results. The hypothesis was incorrect; however, the insignificant amount of change in academic achievement ultimately highlights that the resources being put into free school lunches are not enough to benefit students' performance. Instead, more research needs to be done on the free school lunches to produce noticeable change in academic results.

KEYWORDS: Behavioral and Social Sciences, Sociology and Social Psychology, Free School Lunches, Student Health Academic Performance.

Introduction

Food, water, and sleep are all essential to the survival of humans, but many do not have the resources to be able to have all three at once. Of course, everyone needs these resources, but children and adolescents need them to keep up with school. However, not all children are able to have meals at home, as 18 % of students have moderate to severe food deprivation, with one-third of inner city public school kids not eating enough calories, which is where universal school lunches come into play.[1] Free school meals are the one time a day when children are guaranteed a meal to keep them fulfilled during the day.[2] Even for food-secure families, free school meals provide an opportunity for students to eat a supposedly healthy meal without having to pay for the meal. School meals have a direct impact on the growth and health of students, as that is one of the meals they eat every day.[3] In turn, brain development is aided by caloric intake during the learning period.[4] Without food, students' brains won't be able to function properly, and they will not benefit from the time spent in class.

In the past, school meals tended to be very unhealthy with lots of saturated fats and sugars, and a lack of healthier fruits, vegetables, and grains. Consequently, obesity increased in the child population, and many children were still not receiving the proper foods to stimulate brain function.[5] This combination led to criticism of school lunches for being costly to the government while still barely providing children with nutrition.[6] Schools began to institute healthier lunches to improve the health of students. In 2023, $15 billion was allocated to improving the health of school lunches in California.[7] Frozen pizzas and chicken nuggets were replaced with organic options, salads, free-range grilled chicken, and many other foods low in saturated fats, preservatives, and sugars. School lunch meals were turning into gourmet meals, with executive chefs, new equipment, and contracts with local organic farms all being implemented into California public schools. Access to fruits and vegetables for students affects the school environment, which, in turn, affects the students' health, well-being, and focus.[8] Students who ate meals at school had better attendance, higher academic performance, and improved overall health.[7] Attendance also affects student performance, as students who miss school are more likely to miss out on important lectures. When absenteeism is raised, test scores should be lower because students don't get the face-to-face learning experience necessary to fully understand topics.[9] Learning information on a student's own time is much more difficult than being present in class and hearing the teacher speak. Consequently, students with high absenteeism rates will be less likely to perform well on exams and homework. Just before the 2022-23 school year began, California instituted free school lunches for all students in California public schools, one of the first states to do so.[10] However, prior to that, school meal programs were means-tested. For example, students with family incomes below 130% of the federal poverty line received free meals from their schools, but all other students either received reduced-price meals (below 185% of the federal poverty line) or had to bring their own meals for lunch.[6]

DOI: 10.36838/v7i10.133

Different from prior literature, this research paper focuses on the change once universal school lunches were introduced. Prior literature solely focused on the school lunches' nutritional content or the time of consumption as a way to see changes in academic performance. Also, because of the fairly new implementation of universal free lunches, which not all states in the United States have, there has not been much research on the topic of free school meals. The goal of this research is to create a cause for change regarding other states and the implementation of universal school meals because many children do not get nutritious food at home, and rely on school lunches as their primary source of healthy food. This research is meant to benefit the students who currently deal with malnutrition, which impacts their ability to learn and focus in class. This can ultimately correlate to the loss of opportunities in life, including job opportunities. Food is essential to learning, and food is essential to life, so withholding it from students in need can cause issues that affect people for their entire lives.

This research paper asks if receiving universal free lunches increases academic performance in students. It is hypothesized that universal school lunches will impart a positive change on students' academic performance and a decrease in the rate of absenteeism.

Methods

First, all academic record files were downloaded from the California Assessments of Student Performance and Progress (CAASPP) web page in the assessments tab, and this used the English Language Arts/Literacy and Mathematics (ELPAC) exam.[11] All absenteeism files were downloaded from the data and statistics tab on the California Department of Education Website.[12] The data was then downloaded into RStudio code using readr, a tidyverse package, R version 4.4.2 (Pile of Leaves), and manipulated from there using other tidyverse packages. All the files were harmonized, and schools with unknown values for any important columns were disregarded. Grades 3, 4, 5, 6, 7, 8, and 11 were included in this study. Certain schools were also exempt from the experiment because some schools did not report data for every year that was included in this study. All years from the 2014-2015 school year to the 2023-2024 school year were included, except for the school years of 2019-2020 and 2020-2021, due to a lack of available data from circumstances resulting from the COVID-19 pandemic for academic performance. Absenteeism was compared from 2018-19 to 2023-24. To harmonize the files with the same data, but different titles for the set of data, the titles were mutated to become the same name, which occurred in the case of "Test ID" and "Test Id." For the data to be put together, their titles had to be the same, so they were mutated to both be named "Test ID" using dplyr. The academic performance data was then placed into a Wilcoxon test to determine if the data showed a shift between years. RStudio produced graphs, plots, and other visualizations using ggplot2. Chronic absenteeism was calculated into two different plots to compare the different years side by side using ggplot2. Data normalization occurred to help organize the data, which was cleaned prior to analysis.

Result and Discussion

Results of Academic Performance:

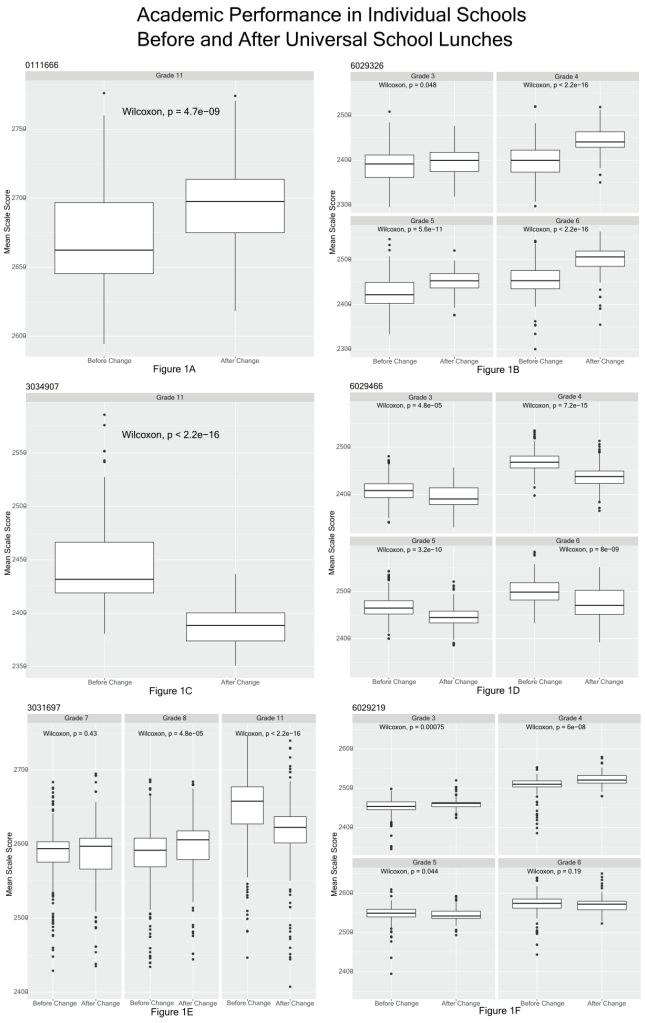


Figure 1: The school code in the top left corner represents an individual school in the Newport Mesa Unified School District. The comparison is between before universal school lunches were instituted (left) and after they were instituted (right). A box and whisker plot was used, and p-values less than 0.05 reject the null hypothesis. Mean Scale Score (y-axis) is the average score of all the students in a grade on a scale from 2000 to 3000. Results varied between the six different schools shown, as two schools had an increase in scores, two schools had an increase in scores, and two schools had variation between grades.

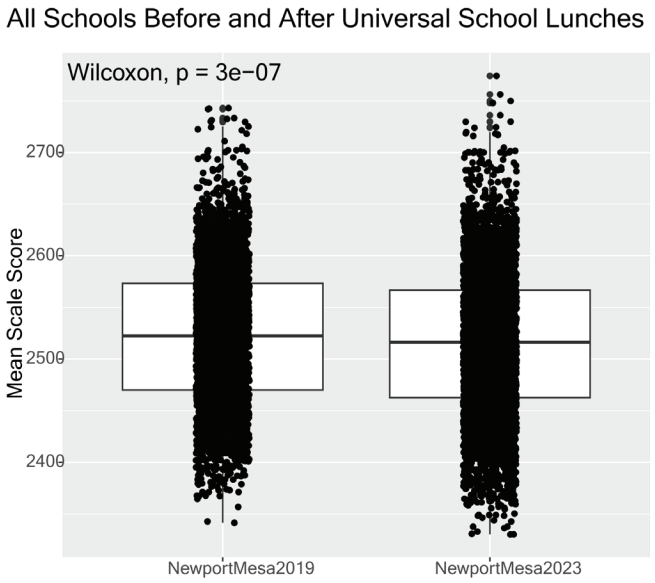


Figure 2: All grades in all schools are in this box and whisker plot, and each dot represents an individual attending a school in the Newport Mesa Unified School District. The comparison is between before universal school lunches were instituted (left) and after they were instituted (right). A box and whisker plot was used, and a p-value of less than 0.05 rejects the null hypothesis. Mean Scale Score (y-axis) is the average score of all the students in a grade on a scale from 2000 to 3000. From 2019 to 2023, the average score for the entire school district decreased, but not to a significant extent.

From the data in the score comparison from before and after the change in having school lunches, there are varying results for each school. In Figure 1A, there is a significant increase from before the change to after the change, as the difference in average mean scale score after the change is significant when compared to before the change. In Figure 1B, there is an increase in mean scale score for all grades shown, while in Figure 1C, there is a decrease in the grade shown. In Figure 1D, there is a significant decrease in the mean scale score for each grade, especially the fourth grade. In Figure 1F, some grades increased, some decreased, and some remained the same.

Both increasing scores and decreasing scores are observed in the many different schools tested. The schools in Figures 1A and 1B all increased, while schools in Figures 1C and 1D decreased, and Figures 1E and 1F had variation between grades in the school. This highlights the discrepancies between the different schools, as there is no pattern shown. Even grades within the same school had opposite results, as in Figure 1E, where grade 8 had scores increase while grade 11 had scores decrease. Even within schools, different grades had different magnitudes and directions of score changes. The other grade with a statistically insignificant p value is grade 6 in Figure 1F. Though it seems as if the points on the plot look much higher after the change, outliers make the mean of before and after the change nearly identical.

In Figure 2, there is a decrease in the mean scale score shown. The mean scale scores in 2019 had a smaller range in comparison to 2023, but 2023 had more outlier scores. These outliers increased the range for the 2023 plot. The 2023 plot had mean scores reach above 2750, but 2019 did not. At the same time, 2019 only had two below 2350, while 2023 had more than ten below 2350, highlighting the larger range of the 2023 plot.

A big issue faced during this study was the lack of available data for some schools during certain years, other than the pandemic years. Nearly five schools' data were unusable because of gaps in data during certain years, especially early in the year range, like 2014 and 2015.

Results of Absenteeism:

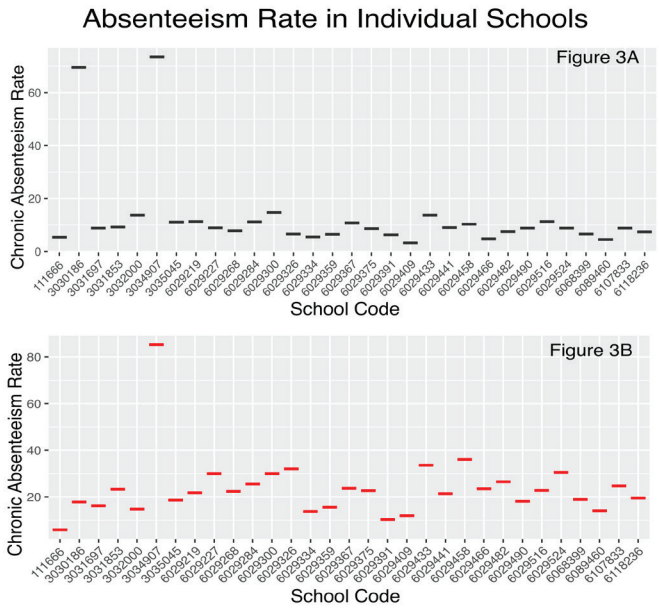


Figure 3: The students who miss more than 10 percent of the academic year are considered chronically absent and are included in this figure. School code represents each school in the Newport Mesa Unified School District. The school year 2018-19 has the black dashes and is Figure 3A, and 2022-23 has the red dashes and is Figure 3B. There was a significant increase in chronic absenteeism from 2018-19 to 2022-23.

As shown, absenteeism rates in 2018-19 were much lower than in 2022-23. Nearly every school had an absenteeism rate increase from 2018-19 to 2022-23. The main exception is school 3030186, which decreased. The average between all the schools in 2018-19 is closer to 10 percent, whereas in 2022-23, the mean chronic absenteeism rate is larger than 20 percent.

Discussion:

The initial hypothesis was that if students are eating at least one fulfilling meal every day, their academic performance should slightly increase, and absenteeism rates would fall. There is a slight increase in mean scale score, which indicates student performance; however, the increase is not statistically significant because the increase remained within the 75th percentile from the pre-change grouping. The increase is minimal and could more easily be attributed to other factors, such as curriculum changes, different types of students attending the schools, and the pandemic. Therefore, the null hypothesis cannot be rejected, highlighting the need for further research to be done on this topic to encourage a proven answer to this hypothesis. The reason for the large increase in absenteeism is the obvious standout of the pandemic. After the pandemic,

DOI: 10.36838/v7i10.133

many students began to skip school more because of illness, but also because of a lack of motivation caused by online schooling. Many students skipped online classes because of the lack of knowledge by teachers about technology, allowing students to take advantage of that. One in four students in the 2022-23 school year in the US were chronically absent.[13] Students didn't feel motivated to go to school because they were used to sitting at home all day, staring at their computer screens. The other reason is that students of lower socioeconomic status had to help out their families at home during the pandemic, and never were able to relinquish that responsibility after the pandemic ended.

This study was necessary because it is important to highlight whether the school lunches children are receiving are worth the money that they cost the government. Instead, parents packing their children's lunches could be more cost-efficient for the government, though it clearly would not be accessible for all families to supply their children with home-cooked meals every day. This introduces the question of whether having fast food packed by parents in a rush to get to work is healthier than school-provided lunches. Could the money already allocated to free school lunches be spent on providing more nutritious meals for children? Though in the Newport Mesa Unified School District, 48% of daily fruit, 40% of daily dairy, and 30% of daily grain are included in student school lunches, the amount of salt and preservatives used to keep the lunches fresh is unknown to the public, requiring further research.[14] It is also important to highlight that students need hearty lunches to get them through the day, but if they aren't receiving lunches, then they won't be able to focus in school. Certain foods, like fish, unrefined cereals, and whole grains, promote brain function and increase cognitive ability; whereas refined cereals, grains, red meat, and unsaturated fats all decrease cognitive functions.[15] Of course, not everything is going to benefit all parts of the body, but eating the same unhealthy foods every day, which is not in moderation, will lead to negative outcomes for an individual's health.

As previously mentioned, some years had to be excluded from the pool of years selected. Namely, the years 2013-14 and all years prior, and 2020-21, the years during the pandemic, had no information gathered due to multiple circumstances. Before 2013-14, there were no records on academic performance; even if there were, the amount of change in schools in a time period longer than a decade causes many uncontrollable variables in the experiment. The impact of the pandemic also cannot be understated. Many students lost valuable time in school due to online learning, and curricula changed consequently to ensure students did not miss out on learning. Schools were set back, and needed to change to help students learn valuable subjects they missed out on.

Conclusion

The topic of school lunches is critical to research because of the impact they have on the developing bodies of children. When children don't eat or eat unhealthy, they, no doubt, are more susceptible to illness and malnutrition, which heavily influence the way their bodies and organs grow. The children of today are the generation that will be in charge of the world within a few decades, and not permitting them to reach their full potential because they did not get enough nutrition to benefit their studies when they were younger is unjust and worrying for the future of our planet. Without being able to fully comprehend the information taught during adolescence, the leaders of tomorrow will not be able to replace the leaders of today. It is also important because many students get one meal a day: their school lunch. Without these lunches, many students will become malnourished.

The shift from no school lunches being provided to them being provided did not make a difference in the Newport Mesa Unified School District. However, that school district lies within Orange County, a county with one of the highest median incomes in the United States at $109,361.[16] If another county's data set were used, likely, the results may not be the same as the results for this county. Higher socio-economic status families will be more able to provide their children with healthier food at home to aid the students' academic success, but in lower-income areas, less expensive foods, which tend to be more unhealthy, will be unable to help academic success. Similarly, higher-income families will be more able to hire a tutor outside of school to help their child do their schoolwork, which inevitably will increase students' scores. Additionally, the Newport Mesa Unified School District was chosen to try to control socioeconomic status because of the higher median income in its county, as fewer people would have been under the line where they received free school meals before they became universal. Another possible model could focus on the impact of individual students before and after universal school lunches were implemented. It would be necessary to account for the expected growth as students' age. Ultimately, socioeconomic status deeply affects research in the behavioral and social sciences because it is uncontrollable.

More research should be done on this topic, including a deeper look into the nutritional values of the school lunches provided to students. If some schools have different nutritional requirements, then their test scores could be compared to see if one is more beneficial to students than another. Also, measuring the impact of eating free school lunches on behavior in schools could be effective. Has the amount of suspensions decreased since the implementation of free school lunches? Additionally, would the results of this experiment differ if other counties or other states with free school lunches were tested? Would results be different in other states where they changed to free school lunches before the pandemic?

Acknowledgments

I attest that the ideas, graphics, and writing in this paper are entirely my own. A special thank you to Matthew Montierth and Indigo Research for their guidance while writing this paper.

References

1. Kleinman, R. E.; Hall, S.; Green, H.; Korzec-Ramirez, D.; Patton, K.; Pagano, M. E.; Murphy, J. M. Diet, Breakfast, and Academic

Performance in Children. *Ann. Nutr. Metab*. **2002**, *46* (Suppl. 1), 24–30. https://doi.org/10.1159/000066399.
2. Earl, L.; Lalli, G. S. Healthy Meals, Better Learners? Debating the Focus of School Food Policy in England. *Br. J. Sociol. Educ.* **2020**, *41* (4), 476–489. https://doi.org/10.1080/01425692.2020.1735999.
3. Lundborg, P.; Rooth, D.-O.; Alex-Petersen, J. Long-Term Effects of Childhood Nutrition: Evidence from a School Lunch Reform. *Rev. Econ. Stud*. **2022**, *89* (2), 876–908. https://doi.org/10.1093/restud/rdab028.
4. López-Gil, J. F.; Mesas, A. E.; Álvarez-Bueno, C.; Pascual-Morena, C.; Saz-Lara, A.; Cavero-Redondo, I. Association Between Eating Habits and Perceived School Performance: A Cross-Sectional Study Among 46,455 Adolescents From 42 Countries. *Front. Nutr.* **2022**, *9*. https://doi.org/10.3389/fnut.2022.797415.
5. Florence, M. D.; Asbridge, M.; Veugelers, P. J. Diet Quality and Academic Performance. *J. Sch. Health* **2008**, *78* (4), 209–215. https://doi.org/10.1111/j.1746-1561.2008.00288.x.
6. Altindag, D. T.; Baek, D.; Lee, H.; Merkle, J. Free Lunch for All? The Impact of Universal School Lunch on Student Misbehavior. *Econ. Educ. Rev.* **2020**, *74*, 101945. https://doi.org/10.1016/j.econedurev.2019.101945.
7. Jones, C. Goodbye Hotdogs, Hello Vegan Masala: California's School Lunches Are Going Gourmet. *CalMatters*. August 10, 2023. http://calmatters.org/education/k-12-education/2023/08/school-lunches/ (accessed 2025-02-24).
8. Johnson, D. B.; Podrabsky, M.; Rocha, A.; Otten, J. J. Effect of the Healthy Hunger-Free Kids Act on the Nutritional Quality of Meals Selected by Students and School Lunch Participation Rates. *JAMA Pediatr*. **2016**, *170* (1), e153918. https://doi.org/10.1001/jamapediatrics.2015.3918.
9. *Student Absences, Instructional Time, and Academic Achievement*. NBER. https://www.nber.org/digest/nov14/student-absences-instructional-time-and-academic-achievement (accessed 2025-02-24).
10. *Universal School Meals Help All California Children Thrive*. California Budget and Policy Center. https://calbudgetcenter.org/resources/universal-school-meals-help-all-california-children-thrive/ (accessed 2025-02-24).
11. *2023–24 Smarter Balanced ELA and Mathematics Test Results at a Glance - CAASPP Reporting (CA Dept of Education)*. https://caaspp-elpac.ets.org/caaspp/DashViewReportSB?ps=true&lstTestYear=2024&lstTestType=B&lstGroup=1&lstSubGroup=1&lstSchoolType=A&lstGrade=13&lstCounty=00&lstDistrict=00000&lstSchool=0000000 (accessed 2025-02-24).
12. *Chronic Absenteeism Data - Accessing Educational Data (CA Dept of Education)*. https://www.cde.ca.gov/ds/ad/filesabd.asp (accessed 2025-02-24).
13. Lurye, J. G., Bianca Vázquez Toness, Sharon. *Millions of kids are still skipping school. Could the answer be recess — and a little cash?*. The Hechinger Report. http://hechingerreport.org/millions-of-kids-are-still-skipping-school-could-the-answer-be-recess-and-a-little-cash/ (accessed 2025-02-24).
14. *Newport-Mesa Unified School District - School Nutrition And Fitness*. https://nmusdnutrition.com/index.php?page=menus&sid=1008152208200351 (accessed 2025-02-24).
15. Puri, S.; Shaheen, M.; Grover, B. Nutrition and Cognitive Health: A Life Course Approach. *Front. Public Health* **2023**, *11*, 1023907. https://doi.org/10.3389/fpubh.2023.1023907.
16. Bureau, U. C. *Median Household Income by County in the United States and Puerto Rico*. Census.gov. https://www.census.gov/library/visualizations/interactive/median-household-income.html (accessed 2025-02-24).

Author

Sascha Pirghibi, a student at Sage Hill High School, plans to study biomedical engineering in college. Sascha is a student-athlete who plays soccer and enjoys learning about biology and mathematics. He cares strongly about his community, peers, and family, and wants to make a difference in the world.

DOI: 10.36838/v7i10.133

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Specific Genetic Factors in PTSD and Comorbidities

Sophia Tang

Boston University Academy, 1st University Rd, Boston, MA, 02021, USA; tsophia200712@gmail.com
Mentor: Collin McColl

ABSTRACT: Post-Traumatic Stress Disorder (PTSD) is a mental health disorder caused by traumatic events. Like many other disorders, it is highly comorbid and, though mostly triggered by traumatic ordeals, there are many different underlying causes beneath its development. One such example is genes, many of which are being researched to this day. This study has identified several genes linked to PTSD and other disorders found to be highly comorbid with PTSD, such as anxiety or depression. This includes dopamine-associated genes like DRD2, SLC6A4, SLC6A3, and those related to the hypothalamic-pituitary-adrenal (HPA), like CRHR1, BDNF, and COMT. Identifying and understanding these genes and many similar genetic pathways allows the formation of precise, individualized treatments such as targeted pharmacological treatments and tailored psychotherapy.
KEYWORDS: Biomedical and Genetics and Molecular Biology of Disease, PTSD, Comorbidity.

Introduction

Post-Traumatic Stress Disorder (PTSD) is a mental health condition triggered by experiencing or witnessing traumatic events such as combat, assault, accidents, or natural disasters. Common risk factors include intense or prolonged trauma, previous trauma, high-risk professions, mental health issues, lack of support, and genetic predispositions. PTSD, like many other mental disorders, has many factors, including environmental ones.[1] Experiencing intense or prolonged trauma, especially over an extended period, increases the likelihood of developing the disorder. If individuals have faced trauma earlier in life, it can make them more vulnerable to future traumatic events, as their emotional resilience may already be affected. Certain professions, such as military personnel or first responders, face a heightened risk of PTSD due to frequent exposure to distressing and life-threatening situations. However, it is implied that around 30-40% of those who have this condition have it due to their genes.[2] These people, in situations where traumatic events happen, are at a higher risk as the traumatic events are more likely to trigger certain genes that lead to PTSD. Pre-existing mental health conditions like anxiety or depression can also intensify the impact of trauma, making it harder to process and recover from these events. Furthermore, if the patient lacks a supportive network of family and friends, this would lead to feelings of isolation and overwhelmedness, with no emotional outlet or resources to help them cope, which would also inhibit recovery.

People with PTSD exhibit abnormal levels of stress hormones, particularly adrenaline, which is typically released during the "fight or flight" response to danger. This response helps dull pain and heightens the senses to respond to threats. However, individuals with PTSD continue to produce high levels of these hormones even when no danger is present, which may contribute to their emotional numbing and hyperarousal.[3] Brain scans of people with PTSD also show changes in emotional processing regions, such as a smaller hippocampus, which is responsible for memory and emotion. This dysfunction could explain why flashbacks, nightmares, and memory issues persist, as the shrunken hippocampus is unable to properly process these traumatic memories. This review will aim to identify several genes and pathways that lead to the diagnosis of PTSD—including its comorbidities—and investigate why they are important to the research of future treatment.

Diagnostics:

PTSD affects individuals across all ages, with a prevalence of approximately 5.6% of people globally developing PTSD after undergoing a traumatic event.[4] Lower socioeconomic status is associated with increased risk, and chronic PTSD (lasting over 12 months) tends to occur after more severe events, with pronounced avoidance behaviors. PTSD is often accompanied by survivor's guilt, particularly after interpersonal trauma, and can manifest as difficulty regulating emotions, impulsive or self-destructive behaviors, feelings of shame, despair, and hopelessness, as well as social withdrawal and impaired relationships.

To be diagnosed with this disorder, the DSM-5 diagnostic criteria have been categorized into 7 different sections in Table 1.

DOI: 10.36838/v7i10.138

Table 1: DSM-5 diagnostic criteria.

Criterion A: Exposure to actual or threatened death, serious injury, or sexual violence in one or more ways:	Direct experience Witnessed Learned about something happening to close family member / friend Repeated / extreme exposure to adverse details of traumatic event
Criterion B: Event persistently re-experienced(memories, dreams, flashbacks) ... or intense / prolonged distress ... or marked physiological reactions	Recurrent, involuntary, and intrusive distressing memories of the event(s). Recurrent distressing dreams related to the event Dissociative reactions(e.g., flashbacks) in which the individual feels or acts as if the traumatic event(s) were recurring Intense or prolonged psychological distress at exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event Marked physiological reactions to internal or external cues that symbolize or resemble an aspect of the traumatic event
Criterion C: Persistent avoidance of stimuli associated with the event(s), beginning after the event(s), as evidenced by one or both of the following:	Avoidance of or efforts to avoid distressing memories, thoughts, or feelings about or closely associated with the traumatic event. Avoidance of or efforts to avoid external reminders(people, places, conversations, activities, objects, situations) that arouse distressing memories, thoughts, or feelings about or closely associated with the traumatic event(s).
Criterion D: Alteration in cognitions and mood(memory problems; negative beliefs / expectations; distorted thoughts about event; negative emotional states; anhedonia; feelings of detachment; inability to feel positive emotions)	Inability to remember an important aspect of the traumatic event Persistent and exaggerated negative beliefs or expectations about oneself, others, or the world Persistent, distorted cognitions about the cause or consequences of the event(s) that lead the individual to blame himself/herself or others Persistent negative emotional state Markedly diminished interest or participation in significant activities Feelings of detachment or estrangement from others Persistent inability to experience positive emotions
Criterion E: Alteration in arousal and reactivity	Irritable behavior and angry outbursts Reckless or self-destructive behavior Hypervigilance Exaggerated startle response Problems with concentration Sleep Disturbance
Criterion F: Symptoms last > one month	
Criterion G: The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning	

Diagnostic criteria of DSM-5. (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) the diagnostic and manual by the American Psychiatric Association. Each criterion, except for F and G, has some general details.[5] This figure provides the common standards for a patient's diagnosis and serves as a reference for PTSD symptoms.

From a cognitive perspective, PTSD may involve classically conditioned fear responses and irrational associations with stimuli, leading individuals to interpret standard situations as threats. Socially, positive coping strategies and support systems can mitigate such symptoms, whereas inconsistent or absent support exacerbates them. Physiologically, heightened neurological arousal increases susceptibility to conditioning, and prolonged stress induces neurochemical changes that worsen or introduce new symptoms. Neuroimaging studies have highlighted the roles of the medial prefrontal cortex—the anterior cingulate, hippocampus, and the amygdala in stress and memory.[6] Treatments typically involve a combination of psychotherapy, such as exposure therapy or cognitive restructuring, and medication like selective serotonin reuptake inhibitors (SSRIs) for the treatment of PTSD symptoms.[1]

PTSD itself can occur in all people, of any ethnicity, nationality, or culture, and at any age. An estimated 1 in 11 people will be diagnosed with PTSD in their lifetime.[7] However, women are twice as likely as men to have PTSD. Three ethnic groups—U.S. Latinos, African Americans, and Native Americans/Alaska Natives—are disproportionately affected and have higher rates of PTSD than non-Latino whites. This data set suggests that genetic risk factors greatly affect the probability of PTSD, and many studies[4-5,6,8-9] have shown that although specific genetic factors haven't been identified, PTSD is highly polygenic, meaning it is associated with thousands of genetic variants throughout the genome, each making a small contribution to the disorder.

Epigenetics:

The modification of genetic effects by other genes or by environmental factors not accounted for across studies may be a cause of the inconsistency in the associations between genetic mutations and PTSD. For example, some studies have found significant effects for specific genetic variants only under conditions of extreme traumatic stress or with a family history of PTSD.[10] In environmental changes, childhood adversity appears to be a particularly potent modifier of genetic risk for PTSD.[11] Social context also seems to modify PTSD risk; an example being the S allele's association with decreased PTSD risk in environments with low crime and unemployment rates, but with increased PTSD risk in the opposite environment. A recent study also showed that SLC6A4 methylation modified the effect of the number of traumatic events on PTSD after controlling for the SLC6A4 genotype.[11]

DNA methylation is one of the major mechanisms of epigenetic regulation. It involves chemical modifications that regulate DNA accessibility, which in turn alter the transcriptional activity of the surrounding loci.[11] Methylation at CpG sites downstream of the 5-HTTLPR region contributes to SLC6A4 expression, and people who have a greater number of experienced traumatic events were found to be at a greater risk for PTSD only if their SLC6A4 methylation levels were low. Conversely, high SLC6A4 methylation levels seem to protect people who have a greater number of traumatic events from developing PTSD. These findings suggest that both genotype and gene-specific methylation patterns contribute to either PTSD vulnerability or to its resilience. Either way, methylation is deeply involved in regulating PTSD diagnosis and prevention. Additionally, with the diversity of the human genome, it is almost impossible to assume that SLC6A4 is the only gene affecting PTSD. Genes involved in stress response, depression, and emotional regulation—all symptoms of PTSD (Table 1)—are also likely to influence PTSD as well, and many of these genes are also involved in other disorders.

Comorbidity, the coexistence of two or more mental disorders in one patient, is very common among patients with psychiatric disorders, with a rate of around 45% of patients satisfying the criteria.[8] Evidence also suggests that PTSD is a subtype of MDD, a major depressive disorder.[9] Common genetic variations significantly contribute to the heritability of brain disorders, suggesting a shared genetic basis for conditions like PTSD, depression, schizophrenia, and psychosis. Analyzing polygenic risk scores has also been associated with PTSD and MDD with similar post-traumatic stress symptoms in veterans,[12] which may suggest that these two share many genetic similarities. Earlier studies have shown that PTSD shares a similar biological pathway to major depressive disorder and other anxiety disorders.[13] Identifying these shared genetic factors could help researchers pinpoint key genes or molecular pathways that are involved in the development of these disorders.

An example of a core pathway is the dopamine system. This molecular pathway is connected with the core synaptic circuits involved in PTSD both structurally and functionally. The role of dopamine in the etiology of PTSD includes increased urinary and plasma levels of dopamine in individuals with PTSD, indicating a significant positive correlation between dopamine levels and severity of PTSD.[14] Dopamine plays a key role in fear memory formation, consolidation, and extinction. Additionally, dopamine is a key component involved in attention, vigilance, arousal, and sleep, which is negatively impacted by PTSD (Table 1). Dysregulations of dopaminergic neurotransmission are involved in the pathophysiological process of PTSD and are closely linked to the symptoms of PTSD patients, such as trauma re-experience, sustained high alertness, emotional numbness, and avoidance behavior.[9] Example genes involved in this hormone include those like DRD2 and SLC6A3, which affect dopamine levels in the brain.

Genes

DRD2:

The dopamine D2 receptor gene (DRD2), located on chromosome 11q, encodes the D2 subtype of the dopamine receptor.[15] This gene has a restriction fragment length polymorphism—differences or variations among people in their DNA sequences at sites recognized by restriction enzymes—called Taq1A(rs1800497) with two alleles named A1 and A2. DRD2's main function is to regulate the synthesis, storage, and release of dopamine, and its mutations can inhibit dopamine production and activity, leading to negative mental effects as well as an increased risk for addiction and neuropsychiatric diseases. Recent studies have also noticed this gene in shaping regulatory aspects of behavior related to approach and reward processing, even dubbing the DRD2 A1 allele as a reward gene.[16] These findings, considered alongside studies of humans and animals that link dopamine D2 receptor binding and DRD2 genotype to depression, anxiety, and impaired social functioning,[17-19] suggest the potential impact of this gene in the development of depressive and anxiety disorders. Therefore, DRD2 genetic variants can be considered as candidate risk factors for PTSD and MDD etiology as well as for prediction, primary prevention, diagnosis, and treatment strategies in the future.[20]

In addition to PTSD, the DRD2 A1 allele was first linked to severe alcoholism over a decade ago, with DRD2 variants being linked to substance abuse, obesity, gambling, and neuropsychiatric disorders such as Tourette's syndrome, PTSD, schizophrenia, and affective disorders.[21] In additional studies with nonclinical subjects, a significant relationship has been reported between dopamine D2 receptors and depression through report and interview methods of target subjects.[22] Supporting this, the DRD2 Taq I A1 allele was found to be highly positively correlated with PTSD as it moderates sensitivity to stress and the expression of emotional disturbance involving PTSD symptoms.[23] In a previously mentioned study using the 12-Item General Health Questionnaire(GHQ-12) to used to detect possible psychological disorders, subjects with the DRD2 A1 allele compared to those without this allele had significantly higher scores on GHQ_2(anxiety/insomnia), GHQ_3(social dysfunction), and GHQ_4(depression).[19] This implies that DRD2 variants are associated with severe comorbid psychopathology in PTSD subjects. Children with an A1 allele had higher scores on the child behavior checklist(CBCL) anxious/depressive subscale, and the research group concluded associations between the DRD2 A1 allele and early-emerging anxious and depressive symptoms in a community sample of preschool-aged children.[22] If these variations can be similarly identified at an early age in the aforementioned studies, it would greatly aid in an individual's prevention of PTSD and other disorders.

A study on another DRD2 polymorphism, 957C>T—a single-nucleotide polymorphism(SNP) within the DRD2 gene, where a cytosine(C) at position 957 can be replaced by a thymine(T)—found that its C allele was associated with a higher risk of PTSD in war veterans.[24] PTSD individuals were more likely to carry the C allele compared to the healthy control group, and their results suggest that the 957C>T polymorphism contributes to the genetic susceptibility to PTSD. Overall, studies lean towards a heavy association between DRD2 and PTSD diagnosis and the disorder's comorbidities, reinforcing the gene's central role in the dopaminergic dysregulation often observed in trauma-related disorders. This highlights the importance of DRD2 as a potential target for both diagnostic screening and the development of future therapeutics aimed at treating PTSD and its associated disorders.

COMT:

The Catechol-O-methyltransferase (COMT) enzyme is involved in the catalysis and inactivation of catecholamines, such as dopamine, which are involved in regulating mood. A popular choice of study is the rs4680 variant within the COMT gene, which substitutes the amino acid valine(Val) for methionine(Met) at codon 158, and is commonly known as the Val158Met polymorphism.[25] Met allele carriers have a 40% reduction in enzyme activity, resulting in higher levels of dopamine in the brain.[26] The Met158 allele is also associated with a decreased ability to extinguish conditioned fear, a key feature of animal models of PTSD. As a result, higher levels of dopamine can potentially be positively correlated with fear extinction.

Genetic association with PTSD was primarily driven by arousal symptoms, as the arousal cluster was the only PTSD symptom cluster associated with Met allele carriers.[27] These findings suggest that COMT Met allele carriers may be at increased risk for an arousal-based presentation of PTSD, whereas Val homozygotes may be at increased risk for a presentation that has more in common with depression. Arousal symptoms are a substantial component of the DSM-5 criteria (Table 1). Children who were carriers of the Met allele were almost seven times more likely to be diagnosed with PTSD than the non-carriers.[27] Another study on veterans found that the interaction between high trauma exposure and COMT was a prediction of PTSD symptoms.[28] Compared to those with either homozygous genotype, those with the heterozygous genotype(Val/Met) showed fewer symptoms associated

DOI: 10.36838/v7i10.138

with trauma exposure, indicating that those with the homozygous genotype have a higher vulnerability to PTSD.

The Val158met polymorphism is also implicated in the treatment response of major depression. In patients with major depression, the COMT 158val/val genotype—associated with higher enzymatic activity and thus reduced dopamine availability—was significantly linked to a poorer response to antidepressant treatment.[29] The genotype-dependent differences in PTSD may align with COMT's influence in depression, where the Val/Val genotype leads to treatment resistance, which implies that this genotype could potentially result in more depression symptomatology in PTSD. Overall, these studies point to COMT's role as a potential biomarker for PTSD, and a patient's treatment plan would possibly benefit from knowing their COMT genotype, as major differences have been found in their reactions to antidepressants.

SLC6A4:

Genetic variants in dopamine transporters and specific polymorphisms in the serotonin transporter gene SLC6A4, a gene involved in PTSD, can lead to reduced serotonin reuptake, with S-allele carriers exhibiting heightened vulnerability to depression and stress. The serotonin transporter 5-HTT is encoded by a single gene, SLC6A4, and the role of this gene is to transport serotonin at synaptic terminals and other neuronal areas and regulate the emotional parts of behavior. Previous studies have also proven the role of SLC6A4 in the etiology of PTSD. A study by Karestan C. Koenen *et al.*[11] resulted in finding that the number of traumatic events was strongly associated with the risk of PTSD, but only at lower methylation levels. At higher methylation levels, individuals with more traumatic events were protected from this disorder. This suggests that high methylation levels at SLC6A4 modify the effects of traumatic events. As traumatic events are a core part of the causes of PTSD, SLC6A4's effects on their impact on the patient offer a possibility for future investigations of the potential for targeting methylation-mediated gene regulation medications.[11] Polymorphisms in SLC6A4 have also been associated with a wide variety of neurological and psychiatric disorders, including increased risk of depression, obsessive-compulsive disorder(OCD), and substance use disorders.[30]

A recent study found that the SLC6A4 SNP rs16965628—a variation in the DNA sequence where a single nucleotide(A, C, G, or T) is different from the typical or reference sequence at a specific location—and the polymorphic region 5-HTTLPR are implicated in neural reactions in response to traumatic reminders and the cognitive control of emotions in PTSD.[31] rs16965628, which is associated with the expression of the serotonin transporter gene, was found to modulate task related ventrolateral prefrontal cortex (PFC) activation—and after trauma exposure, the gene would influence serotonin pathways in the brain, which would in turn alter the function of the brain regions involved in controlling emotions.

Meanwhile, 5-HTTLPR leans toward modulating the activation of the left amygdala, the area responsible for fear learning. A meta-analysis reviewing the correlation between 5-HTTLPR and PTSD found that carrying the 5-HTTLPR SS genotype is possibly a risk factor for PTSD after high trauma exposure.[32] This suggests that the genotype itself can interact with the environmental factors, such as trauma exposure, to influence the risk of PTSD, and as a result, appear in certain populations such as combat veterans or individuals with high trauma exposure. The HTTLPR genotype alone does not strictly indicate PTSD; however, this genotype, if grouped alongside other genes, can help researchers predict the disorder with higher accuracy and result in early prevention or treatment.

SLC6A3:

The SLC6A3 and SLC6A4 genes are members of a class of neurotransmitter transporters for the release, reuptake, and recycling of neurotransmitters in synapses. SLC6A3 and SLC6A4 encode a dopamine transporter and serotonin transporter, respectively. The 9-repeat(9R) allele of SLC6A3 encodes a dopamine transporter, a member of the sodium- and chloride-dependent neurotransmitter transporter family, which plays a key role in the regulation of dopaminergic neurotransmission by removing dopamine from the synaptic cleft via reuptake through the transporter.[33]

The SLC6A3 gene provides instructions to make the protein dopamine transporter (DAT), which is associated with depression and deals with negative emotional stimuli. Although the specific causes for this association have not been found, studies have shown that variations of the SLC6A3 gene have been associated with higher rates of depression and suicidal behavior.[34] The role of dopamine in the etiology of PTSD is also supported by findings of elevated urinary and plasma levels of dopamine among those affected by the disorder, and by reports of a significant correlation between dopamine concentration and severity of PTSD symptoms in affected individuals.[35] Additionally, previous work indicates that 9R allele carriers of the SLC6A3 3′UTR VNTR polymorphism show a significantly increased risk of lifetime PTSD. In contrast, a systematic review has concluded that a correlation between 3′UTR VNTR(rs28363170), a genetic variant of SLC6A3, and MDD is inconclusive, with two studies showing a positive association—that the genetic variant is a risk factor in developing MDD—and two other studies resulting in a negative correlation.[36] Although findings regarding SLC6A3's link to major depressive disorder (MDD) remain mixed, the presence of studies indicating a positive association highlights the gene's potential relevance and underscores the need for further research to clarify its role in the development of mood and stress-related disorders. In conclusion, while the role of the SLC6A3 gene in depression and PTSD is not yet fully understood, emerging evidence suggests that its polymorphisms—particularly the 9R allele of the 3′UTR VNTR variant—may contribute to increased vulnerability to PTSD.

HPA Axis: Cortisol:

In addition to dopamine, PTSD is often associated with alterations in the hypothalamic-pituitary-adrenal (HPA) axis. It consists of the hypothalamus, the brain region's control center, the pituitary gland, and the adrenal glands.[37] The hypothal-

amus releases the corticotropin-releasing hormone (CRH), which stimulates the pituitary gland to release adrenocorticotropic hormones (ACTH), which then causes the adrenal glands to release cortisol. It is one of the major stress response systems regulating the secretion of the glucocorticoid cortisol, as cortisol helps the human body to regulate mood and behavior.

Dysregulation of the HPA axis alters cortisol feedback and morning cortisol release in PTSD patients, causing enhanced negative feedback to the HPA axis—this is abnormalities in cortisol levels were among the first findings in individuals with PTSD. Under controlled conditions, lower circulating levels of cortisol have repeatedly been described in subjects with PTSD in comparison to healthy individuals.[38] Defined as hypocortisolism, the deficiency of cortisol in PTSD has been explained through chronic stress models which propose that HPA axis activity is increased in early stages, but decreases with a longer duration of stress, and subsequently reaches a state of hypocortisolism, which has been suggested as a key component in the development of various stress-related disorders. The fact that cortisol deficit may trigger PTSD is also supported by findings that the administration of supplemental doses of cortisol to acutely ill medical patients reduces the PTSD outcome.[39] Elevated glucocorticoid levels were found to inhibit memory retrieval, and in clinical trials, the administration of cortisol provided significant positive effects on memory-related PTSD symptoms.

Enhanced negative feedback in the HPA axis can also cause depression and increased stress as the HPA axis becomes overly sensitive to cortisol. Similarly, HPA axis overactivity has also been identified in major depression, and a study associated patients with major depression with psychosis(PMD) with higher cortisol, in addition to a general higher elevation of HPA axis activity positively correlated to specific depressive subtypes that include psychosis.[40] Involved in many disorders, cortisol has been shown to give positive treatment effects to PTSD, and as both the disorder and the hormone are deeply involved in MDD, this gives potential for treatment strategies targeting cortisol regulation to be beneficial for managing symptoms of both disorders simultaneously.

HPA Axis: CRHR1:

A gene connected to the HPA Axis is the CRHR1 gene. The CRHR1 gene encodes the corticotropin-releasing hormone receptor type I(CRHR1) gene, a protein that is essential for the activation of signal transduction pathways that activate mesolimbic and HPA axis responses to many types of stress.[41] CRH is one of the key stress factors in the central nervous system. The CRHR1 receptor is crucial for establishing the initial HPA axis response to stressful events. Variation in the CRHR1 gene has also been shown to moderate the relationship between childhood maltreatment and cortisol response to the Dex/CRH test.[42]

Several other studies have also linked CRHR1 polymorphisms to MDD in interaction with childhood maltreatment or negative experiences.[43-46] In relation to PTSD, numerous studies have investigated CRHR1 variation and depression in the context of a traumatic event. For example, a study showed that the rs110402 polymorphism was associated with cortisol levels and the development of depression in adulthood in males who experienced childhood trauma.[47] Another study found that the same polymorphism regulated the relationship between child abuse and adult depression,[43] with the effect of rs110402 consistent with that reported by the previous study. This suggests that CRHR1 variations have the potential to be a biomarker for depression.

A protective effect of the CRHR1 TAT haplotype (rs7209436, rs110402, rs242924) was also discovered,[43] which was later replicated in a study by Polanczyk *et al.*[46] examining adult depression following childhood maltreatment. However, a different polymorphism, rs12944712, was found to be significantly related to acute PTSD symptoms.[48] A study tested this hypothesis in an epidemiologic sample of adults exposed to one of the 2004 Florida hurricanes and found that polymorphisms in the CRHR1 gene are related to post-hurricane PTSD symptoms and rs12938031 PTSD diagnosis in adults.[49] Overall, evidence suggests that CRHR1 variations may be important in PTSD, as well as other MDDs.

BDNF:

The brain-derived neurotrophic factor (BDNF) also plays a crucial role in the development of PTSD, as it is linked to depression and the extinction of fear, which is impaired in PTSD. BDNF regulates neuronal survival, growth, differentiation, and synapse formation, and is known to be associated with MDD, anxiety disorders, and post-traumatic stress disorder (PTSD).[50]

BDNF is associated with PTSD risk and the exaggerated startle reaction (Table 1) in United States military service members who were deployed during the wars in Iraq and Afghanistan.[51] A common single-nucleotide polymorphism (SNP) in the BDNF gene leading to a valine to methionine substitution at position 66 influences human hippocampal volume, memory, and susceptibility to PTSD. Recently, a study has demonstrated that the frequency distribution of the Val-66Met polymorphism was different between subjects with and without PTSD.[51] The frequencies of the Met/Met genotype and Met carriers are significantly higher in individuals with PTSD than those without PTSD. These findings are also consistent with other results showing that the frequency of the Met/Met was greater among those with PTSD than non-PTSD controls. Another study has also shown that the BDNF Val66Met genotype predicts response to cognitive behavior therapy in PTSD and possesses evidence that BDNF facilitates extinction learning.[52] Additionally, another study used clinical assessments childhood trauma questionnaire(CTQ), and the posttraumatic stress disorder checklist(PCL) to find that people with the rs6265 Val/Val genotype and higher CTQ scores have higher PCL tests.[53] This study associated the interaction between CTQ and the BDNF polymorphism to be significant to PTSD symptoms, in addition to being positively correlated to the thickness of the left fusiform gyrus(IFFG) and the transverse temporal gyrus. The same thickness of both regions is also heavily connected to symptoms like depression

DOI: 10.36838/v7i10.138

and anxiety. Therefore, finding a way to moderate these genes may impact the direction of future research on treating the symptoms of PTSD and its comorbidities.

BDNF has also been shown to be connected to serotonin, linking it to MDD. Serotonin pathways influence BDNF expression, while BDNF regulates serotonin receptor development, both of which contribute to mood disorders like depression[54] and trauma stressor-related disorders like PTSD. With current research like this increasing, there is more potential in finding more genetic and hormonal connections between the disorders for the purpose of looking into targeted gene therapy.

Treatment

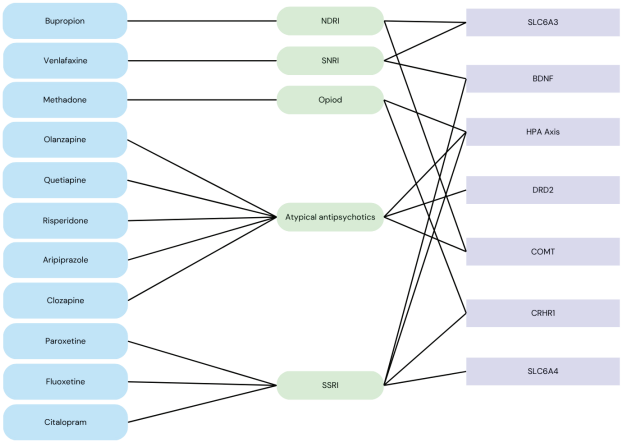
Due to the lack of knowledge about the genetic factors for PTSD, the main treatment is talk therapies with a combination of medications for treating symptoms. Talking therapies include cognitive behavioral therapy (CBT), which aims to help you manage problems by changing how you think and act.[55] The therapist will help the patient gain control of their fear and distress by reviewing any conclusions they have drawn about the experience. Another type of treatment is eye movement desensitization and reprocessing (EMDR), which involves recalling the traumatic incident in detail while making eye movements, usually by following the movement of your therapist's finger.

As of now, the only authorized medications by the FDA for PTSD are selective serotonin reuptake inhibitors (SSRIs): sertraline (Zoloft), paroxetine (Paxil), and venlafaxine (Effexor), a serotonin–norepinephrine reuptake inhibitor (SNRI).[56] These are all types of antidepressants that work by increasing serotonin levels in the brain—SSRIs target serotonin, while SNRIs target both serotonin and norepinephrine. This medication alone will not cure PTSD, but a rise in serotonin levels can improve symptoms and make people more responsive to other types of treatment, such as CBT. In treatment, a psychiatrist will often switch to different SSRI or SNRI medications based on patient response, tolerability, or other issues of slow or rapid metabolism of those particular medications.

However, researching deeper into the genetic causes of PTSD may lead to more medications specialized in treating PTSD symptoms and causes. The most recent research focuses on the use of trauma-focused therapies such as Cognitive Processing Therapy(CPT), Prolonged Exposure Therapy (PE), Eye Movement Desensitization and Restructuring (EMDR), and others with a significant trauma focus.[57] Additional medication could potentially reduce PTSD and its comorbid disorders' symptoms simultaneously; however, benzodiazepines or other sedative-hypnotic medications should be avoided for long-term use due to risks of dependence, cognitive dysfunction, and reduced stress tolerance.

Possibility of Customized Treatments:

Other than benzodiazepines, various other medications could be considered alongside PTSD therapy. A study investigated drug-target interactions associated with PTSD based on a PTSD-related gene set and yielded a comprehensive network that provides insights into the relationship between various drugs and the genes implicated in PTSD.[58] Each drug correlates with a specific set of genes, indicating potential interactions that may influence treatment response and efficacy in affected individuals. Notably, the study's use of gene ontology analysis resulted in finding the top PTSD-associated drugs—including quetiapine, venlafaxine, fluoxetine, bupropion, olanzapine, citalopram, antipsychotics, paroxetine, risperidone, opioids, methadone, aripiprazole, SSRIs, clozapine, and antidepressants—revealing significant links to genes involved in neurotransmitter receptor activity, serotonin binding, and neural processes related to fear learning and extinction, aligning with known mechanisms of PTSD. As these drugs were identified mainly using statistics and functional analysis, clinical trials are required to validate their efficacy and effects, including the drugs already in use. The identified drugs can be categorized into SSRIs, SNRIs, norepinephrine-dopamine reuptake inhibitors (NDRIs), atypical antipsychotics, and opioids. Their categories and the genes they can target directly can be found in Table 2 below. However, the actual treatment effects for each gene require more research and clinical trials to confirm their therapeutic results.



Blue squares include the drug names connected to their respective category (green). The categories are connected to the aforementioned genes (purple), and being connected indicates the drug category being able to directly target the gene.[59-68] This diagram organizes the identified drugs into categories based on their effects, and also shows which genes current research has related these drugs to.

Medical treatments often target gene families rather than single genes because diseases arise from complex genetic pathways involving multiple interacting genes. A drug designed for a specific pathway might be highly effective for one disease but less so for another, depending on how that pathway contributes to each condition. For example, dopamine-related treatments may work well for both Parkinson's disease and schizophrenia, but their effectiveness varies because dopamine dysfunction affects these disorders differently. With future research, individuals with genetic mutations like DRD2 A1, linked to neuropsychiatric disorders like PTSD, have the possibility of taking preventive measures such as avoiding dopamine-stimulating substances like alcohol, engaging

in cognitive training, following a dopamine-balanced diet, and undergoing regular screenings to detect early signs of illness. If symptoms develop, treatment could follow a two-step approach, beginning with general drug therapy to stabilize symptoms, followed by personalized psychotherapy[58] that addresses the specific ways their genetic traits influence behavior. For example, addiction therapy could target dopamine-driven impulsivity, while CBT for depression could focus on dopamine receptor dysfunction's impact on motivation and mood. This gene-pathway-based approach combines targeted drug therapy with talk therapy, which aligns with the mainstream way of treatment and, with further research, has the potential to optimize results.

Current research:

Current research for new treatment includes psychedelics, such as Methylenedioxymethamphetamine (MDMA), psilocybin, and ibogaine.[69] In addition to psychedelic-assisted therapy for PTSD, there is a procedure called stellate ganglion block or SGB.[70] A stellate ganglion block (SGB) is an injection of anesthetic medication into a collection of nerves called the stellate ganglion. These nerves are located in your neck, on both sides of your voice box. The injection can help relieve pain in your head, neck, upper arm, and chest. The stellate ganglion is like a routing center in the nervous system for the fight or flight nervous system, and studies have found that briefly anesthetizing the area effectively treats PTSD symptoms. The shot has some short-term side effects, such as a droopy or reddened right eye, but these diminish after 6 to 12 hours, while the improvements in the PTSD symptoms last longer. While some see improvement with one treatment, others need additional injections. Compared to other methods, SGB has shown positive results in three out of every four patients with PTSD. This treatment effectively resets the body and brain to a pre-trauma state, providing prompt relief and relaxation.

■ Discussion

There has been much new and updated research into PTSD and its treatment over the last few years. Out of the many treatments researched, the trauma-focused therapies, delivered individually by the therapist, showed the most benefit in treatment. There is strong evidence that many of the systems that mediate stress responses also facilitate the encoding of aversive memories, which could form the basis for the development of PTSD[71] and open up avenues for the development of novel prevention strategies. The diversity of these symptoms, like those mentioned in Table 1, suggests the involvement of multiple neurobiological systems. The disorder also has a high degree of comorbidity with other psychiatric disorders such as depression, substance abuse, and panic disorder.

Early intervention in neurological or psychiatric disorders is likely to have a significant impact on outcomes. As seen from the previously mentioned studies in the genetic section, identifying the genetic causes of PTSD is crucial for developing more effective treatments and prevention strategies by addressing individual differences in risk factors, cell structure, and treatment response. Finding common genes between disorders would allow researchers to link treatment plans together. This study has identified several genes linked to PTSD susceptibility, and understanding these genetic influences could allow for early intervention in high-risk individuals. For example, those with a known genetic predisposition to PTSD could benefit from preventive measures like resilience training, stress management techniques, or early therapy after trauma exposure to reduce the likelihood of developing full-blown symptoms.

Additionally, genetic insights could lead to pharmacological treatments by shifting PTSD management toward precision medicine, rather than relying on one-size-fits-all approaches like SSRIs and SNRIs, treatments could be tailored to an individual's specific neurotransmitter imbalances. DRD2 regulates dopamine activity, affecting motivation, reward, and emotional processing, which links it to stress sensitivity, reward dysregulation, and comorbidity with anxiety, depression, and PTSD.[16] As a result, targeting DRD2 pathways in drug development can help address mood dysregulation and other trauma-related symptoms. If dopamine-related DRD2 variants contribute to PTSD by affecting emotional regulation and reward processing,[16] targeted therapies that modulate dopamine signaling could be particularly effective for these patients. Those with a heightened fear response due to genetic factors might also benefit more from exposure therapy, while individuals with dopamine-related emotional dysregulation could receive CBT focused on impulse control and mood stabilization. Knowing which genetic factors influence fear response and memory processing could potentially help refine personalized psychotherapy approaches, ensuring that individuals receive the most effective therapy for their specific neurobiological profile.

■ Conclusion

As the genes and hormones, DRD2, COMT, SLC6A4, SLC6A3, cortisol, CRHR1, and BDNF are highly related to other disorders such as MDD in addition to PTSD, patients may benefit if genes related to comorbidity are identified and included in the personalized treatments. This research is especially relevant for military personnel, first responders, and others in high-risk professions, where genetic screening could help identify those who may need specialized resilience training or targeted interventions before trauma exposure. By integrating genetic research with pharmacology and psychotherapy, PTSD treatment and prevention can become more precise, proactive, and personalized, improving long-term outcomes for those at risk.

■ Acknowledgments

I would like to express my deepest gratitude to my mentor, Collin McColl, for his invaluable guidance, support, and encouragement throughout the process of writing this paper. I would also like to extend my thanks to Indigo Research for their assistance and resources, which played a pivotal role in shaping the final outcome of this work.

I attest that the ideas, graphics, and writing in this paper are entirely my own.

DOI: 10.36838/v7i10.138

References

1. National Institute of Mental Health. Post-traumatic Stress Disorder; National Institute of Mental Health, published 2023. https://www.nimh.nih.gov/health/publications/post-traumatic-stress-disorder-ptsd
2. Banerjee, S. B.; Morrison, F. G.; Ressler, K. J. Genetic Approaches for the Study of PTSD: Advances and Challenges. Neurosci. Lett. 2017, 649, 139– 146. DOI: 10.1016/j.neulet.2017.02.058
3. National Health Service. Causes-Post-traumatic Stress Disorder, National Health Service: U.K., published February 17, 2021. https://www.nhs.uk/mental-health/conditions/post-traumatic-stress-disorder-ptsd/causes/
4. Koenen, K. C.; Ratanatharathorn, A.; Ng. L.; McLaughlin, K. A.; Bromet, E. J.; Stein, D. J.; Karam, E. G: Meron Ruscio, A.; Benjet, C.; Scott, K.; Atwoli,
5. L; Petukhova, M; Lim, C. C. W.; Aguilar-Gaxiola, S.; Al-Hamzawi, A.; Alonso, J; Bunting, B.; Ciutan, M.; de Girolamo, G: Degenhardt, L: Gureje, O: Haro, J. M.; Huang, Y.; Kawakami, N.; Lee, S.; Navarro-Mateu, F.; Pennell, B. E.; Piazza, M; Sampson, N; Ten Have, M.; Torres, Y.; Viana, M. C.; Williams, D.; Xavier, M; Kessler, R. C. Posttraumatic Stress Disorder in the World Mental Health Surveys. Psychol. Med. 2017, 47 (13), 2260- 2274. DOI: 10.1017/S0033291717000708
6. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th ed.; Pearson: Boston, 2013.
7. Bremner, J. D. Neuroimaging in Posttraumatic Stress Disorder and Other Stress-Related Disorders. Neuroimaging Clin. N. Am. 2007, 17 (4), 523-ix. DOI: 10.1016/j.nic.2007.07.003
8. Taylor-Desir, M. What Is Posttraumatic Stress Disorder (PTSD)? American Psychiatric Association, published 2022. https://www.psychiatry.org/patients-families/ptsd/what-is-ptsd
9. van Loo, H. M, Romeijn, J.-W. Psychiatric Comorbidity: Fact or Artifact? Theor. Med. Bioeth. 2015, 36(1), 41-60. DOI: 10.1007/s11017-015-9321-0
10. Zhang, F.; Rao, S.; Cao, H.; Zhang, X.; Wang, Q.; Xu, Y.; Sun, J.; Wang, C.; Chen, J.; Xu, X.; Zhang, N: Tian, L; Yuan, J.; Wang, G.; Cai, L: Xu, M.; Baranova, A. Genetic Evidence Suggests Posttraumatic Stress Disorder as a Subtype of Major Depressive Disorder. Clin. Invest. 2022, 132 (3). e145942. DOI: 10.1172/JC145942
11. Koenen, K. C.; Amstadter, A. B.; Nugent, N. R. Gene-Environment Interaction in Posttraumatic Stress Disorder: An Update. J Trauma. Stress 2009, 22 (5), 416-426. DOI: 10.1002/jts 20435
12. Koenen, K. C.; Uddin, M.; Chang, S.-C.; Aiello, A. E; Wildman, D. E.; Goldmann, E.; Galea, S. SLC6A4 Methylation Modifies the Effect of the Number of Traumatic Events on Risk for Posttraumatic Stress Disorder. Depress. Anxiety 2011, 28 (8), 639-647. DOI: 10.1002/da.20825
13. Campbell-Sills, L; Papini, S.; Norman, S. B.; Choi, K. W; He, F.; Sun, X; Kessler, R. C.; Ursano, R. J.; Jain, S.; Stein, M. B. Associations of Polygenic Risk Scores with Posttraumatic Stress Symptom Trajectories Following Combat Deployment. Psychol. Med. 2023, 53 (14), 6733-6742. DOŁ 10.1017/S0033291723000211
14. Smoller, J. W. The Genetics of Stress-Related Disorders: PTSD, Depression, and Anxiety Disorders. Neuropsychopharmacology 2016, 41(1), 297- 319. DOI: 10.1038/npp.2015.266
15. Drury, S. S; Theall, K. P: Keats, B. J. B.; Scheeringa, M. The Role of the Dopamine Transporter (DAT) in the Development of PTSD in Preschool Children. J. Trauma. Stress 2009, 22 (6), 534-539. DOI: 10.1002/jts 20475
16. Eisenberg, D. T. A.; Mackillop, J; Modi, M.; Beauchemin, J.; Dang, D.; Lisman, S. A.; Lum, J. K; Wilson, D. S. Examining Impulsivity as an Endophenotype Using a Behavioral Approach: A DRD2 Taql A and DRD4 48-bp VNTR Association Study. Behav. Brain Funct. 2007, 3, 2. DOI: 10.1186/1744-9081-3-2
17. Bowirrat, A.; Oscar-Berman, M. Relationship Between Dopaminergic Neurotransmission, Alcoholism, and Reward Deficiency Syndrome. Am. J. Med. Genet. B Neuropsychiatr. Genet. 2005, 1328 (1), 29-37. DOI: 10.1002/ajmg.b.30080
18. Shively, C. A.; Grant, K. A.; Ehrenkaufer, R. L.; Mach, R. H; Nader, M. A. Social Stress, Depression, and Brain Dopamine in Female Cynomolgus Monkeys. Ann. N. Y. Acad. Sci. 1997, 807, 574-577. DOI: 10.1111/j.1749-6632.1997.tb51972.x
19. Schneier, F. R.; Liebowitz, M. R.; Abi-Dargham, A.; Zea-Ponce, Y.; Lin, S. H; Laruelle, M. Low Dopamine D(2) Receptor Binding Potential in Social Phobia. Am. J. Psychiatry 2000, 157 (3), 457-459. DOI: 10.1176/appi.ajp.157.3.457
20. Lawford, B. R.; Young, R.; Noble, E. P.; Kann, B.; Ritchie, T. The D2 Dopamine Receptor (DRD2) Gene Is Associated with Co-morbid Depression, Anxiety and Social Dysfunction in Untreated Veterans with Post-traumatic Stress Disorder. Eur. Psychiatry 2006, 21(3), 180-185. DOI: 10.1016/j.eurpsy.2005.01.006
21. Zhang, X.; Han, Y; Liu, X.; Chen, J.; Yuan, Z.; Wang, Y. Assessment of Genetic Variants in D2 Dopamine Receptor (DRD2) Gene as Risk Factors for Post-traumatic Stress Disorder (PTSD) and Major Depressive Disorder (MDD): A Systematic Review and Meta-analysis. J. Affect. Disord. 2023, 328, 312-323. DOI: 10.1016/j.jad 2023.02.001
22. Noble, E. P. The DRD2 Gene in Psychiatric and Neurological Disorders and Its Phenotypes. Pharmacogenomics 2000, 1 (3), 309–333. DOŁ 10.1517/14622416.1.3.309
23. Hayden, E. P.; Klein, D. N.; Dougherty, L. R: Olino, T. M.; Laptook, R. S.; Dyson, M. W: Bufferd, S. J.; Durbin, C. E.; Sheikh, H. I.; Singh, S. M. The Dopamine D2 Receptor Gene and Depressive and Anxious Symptoms in Childhood: Associations and Evidence for Gene-Environment Correlation and Gene-Environment Interaction. Psychiatr. Genet. 2010, 20(6), 304-310. DOI: 10.1097/YPG.0b013e32833adccb
24. Xiao, Y.; Liu, D; Liu, K.; Wu, C.; Zhang, H; Niu, Y.; Jiang, X. Association of DRD2, 5-HTTLPR, and 5-HTTVNTR Gene Polymorphisms with Posttraumatic Stress Disorder in Tibetan Adolescents: A Case-Control Study. Biol. Res. Nurs. 2019, 21(3), 286-295. DOI: 10.1177/1099800419838325
25. Voisey, J; Swagell, C. D: Hughes, I. P: Morris, C. P.; van Daal, A.; Noble, E. P.; Kann, B.; Heslop, K. A.; Young, R. M.; Lawford, B. R. The DRD2 Gene 957C>T Polymorphism Is Associated with Posttraumatic Stress Disorder in War Veterans. Depress. Anxiety 2009, 26(1), 28-33. DOI: 10.1002/da.20517
26. Qayyum, A.; Zai, C. C.; Hirata, Y.; Tiwari, A. K.; Cheema, S.; Nowrouzi, B.; Beitchman, J. H; Kennedy, J. L. The Role of the Catechol-O-Methyltransferase (COMT) Gene Val158Met in Aggressive Behavior, a Review of Genetic Studies. Curr. Neuropharmacol. 2015, 13(6), 802-814. DOI: 10.2174/1570159x13666150612225836
27. Danzi, B. A.; La Greca, A. M. Genetic Pathways to Posttraumatic Stress Disorder and Depression in Children: Investigation of Catechol-O-Methyltransferase (COMT) Val158Met Using Different PTSD Diagnostic Models. J Psychiatr. Res. 2018, 102, 81-86. https://www.sciencedirect.com/science/article/pii/S002239561731316X. DOI: 10.1016/j.jpsychires.2018.03.014
28. Clark, R.; DeYoung. C. G.; Sponheim, S. R.; Bender, T. L.; Polusny, M. A.; Erbes, C. R. ; Arbisi, P. A. Predicting Post-traumatic Stress Disorder in Veterans: Interaction of Traumatic Load with COMT Gene Variation. J. Psychiatr. Res. 2013, 47(12), 1849-1856. DOI: 10.1016/j.jpsychires.2013.08.013

29. Baune, B. T.; Hohoff, C.; Berger, K.; Neumann, A.; Mortensen, S.; Roehrs, T.; Deckert, J; Arolt, V.; Domschke, K. Association of the COMT val158met Variant with Antidepressant Treatment Response in Major Depression. Neuropsychopharmacology 2008, 33(4), 924-932. DOE 10.1038/sj.npp.1301462
30. Miozzo, R.; Eaton, W. W.; Joseph Bienvenu, O. 3rd; Samuels, J.; Nestadt, G. The Serotonin Transporter Gene Polymorphism (SLC6A4) and Risk for Psychiatric Morbidity and Comorbidity in the Baltimore ECA Follow-Up Study. Compr. Psychiatry 2020, 102, 152199. DOI: 10.1016/j.comppsych.2020.152199
31. Morey, R. A.; Hariri, A. R.; Gold, A. L.; Hauser, M. A.; Munger, H. J; Dolcos, F.; McCarthy, G. Serotonin Transporter Gene Polymorphisms and Brain Function During Emotional Distraction from Cognitive Processing in Posttraumatic Stress Disorder. BMC Psychiatry 2011, 11, 76. DOI: 10.1186/1471-244X-11-76
32. Gressier, F.; Calati, R.; Balestri, M.; Marsano, A.; Alberti, S.; Antypa, N.; Serretti, A. The 5-HTTLPR Polymorphism and Posttraumatic Stress Disorder: A Meta-analysis. J Trauma. Stress 2013, 26(6), 645-653. DOI: 10.1002/jts.21855
33. Bannon, M. J.; Michelhaugh, S. K.; Wang, J.; Sacchetti, P. The Human Dopamine Transporter Gene: Gene Organization, Transcriptional Regulation, and Potential Involvement in Neuropsychiatric Disorders. Eur. Neuropsychopharmacol. 2001, 11 (6), 449–455. https://www.sciencedirect.com/science/article/pii/S0924977X01001225. DOI: 10.1016/S0924-977X(01)00122-5
34. Rafikova, E.; Shadrina, M.; Slominsky, P.; Guekht, A.; Ryskov, A.; Shibalev, D.; Vasilyev, V. SLC6A3 (DAT1) as a Novel Candidate Biomarker Gene for Suicidal Behavior. Genes (Basel) 2021, 12(6), 861. DOI: 10.3390/genes12060861
35. Chang, S.-C.; Koenen, K. C.; Galea, S.; Aiello, A. E.; Soliven, R.; Wildman, D. E.; Uddin, M. Molecular Variation at the SLC6A3 Locus Predicts Lifetime Risk of PTSD in the Detroit Neighborhood Health Study. PLOS One 2012, 7(6), e39184. DOI: 10.1371/journal.pone.0039184
36. Gontijo, B. R.; Possatti, L; Fratelli, C. F; Pereira, A. S. R.; Bonasser, L. S. S.; de Souza Silva, C. M.; Rodrigues da Silva, L. C. The 3'UTR VNTR SLC6A3 Genetic Variant and Major Depressive Disorder: A Systematic Review. Biomedicines 2023, 11 (8), 2270. DOI: 10.3390/biomedicines11082270
37. Cleveland Clinic. HPA Axis: The Stress Response System; Cleveland Clinic, published April 12, 2024. https://my.clevelandclinic.org/health/body/hypothalamic-pituitary-adrenal-hpa-axis
38. Schumacher, S; Niemeyer, H.; Engel, S.; Cwik, J. C.; Laufer, S.; Klusmann, H; Knaevelsrud, C. HPA Axis Regulation in Posttraumatic Stress Disorder: A Meta-analysis Focusing on Potential Moderators. Neurosci. Biobehav Rev. 2019, 100, 35-57. DOI: 10.1016/j.neubiorev.2019.02.005
39. de Quervain, D. J.-F. Glucocorticoid-Induced Inhibition of Memory Retrieval: Implications for Posttraumatic Stress Disorder. Ann. N. Y. Acad. Sci. 2006, 1071, 216-220. DOI: 10.1196/annals.1364.016
40. Keller, J.; Gomez, R.; Williams, G.; Lembke, A.; Lazzeroni, L.; Murphy, G. M.; Schatzberg, A. F. HPA Axis in Major Depression: Cortisol, Clinical Symptomatology and Genetic Variation Predict Cognition. Mol. Psychiatry 2017, 22 (4), 527-536. DOI: 10.1038/mp.2016.120
41. CRHR1 Corticotropin Releasing Hormone Receptor 1 [Homo sapiens (Human)] - Gene. Nih.gov, National Center for Biotechnology Information, published 2025. https://www.ncbi.nlm.nih.gov/gene?Db-gene&Cmd=DetailsSearch&Term=1394 (accessed June 11, 2025)
42. Tyrka, A. R.: Price, L. H.; Gelernter, J.; Schepker, C.; Anderson, G. M.; Carpenter, L. L. Interaction of Childhood Maltreatment with the Corticotropin-Releasing Hormone Receptor Gene: Effects on Hypothalamic-Pituitary-Adrenal Axis Reactivity. Biol Psychiatry 2009, 66 (7), 681-685. DOL 10.1016/j.biopsych.2009.05.012
43. Bradley, R. G.; Binder, E. B.; Epstein, M. P.; Tang, Y.; Nair, H. P.; Liu, W.; Gillespie, C. F.; Berg, T; Evces, M.; Newport, D. J.; Stowe, Z. N.; Heim, C. M.; Nemeroff, C. B.; Schwartz, A.; Cubells, JF; Ressler, K. J. Influence of Child Abuse on Adult Depression: Moderation by the Corticotropin-Releasing Hormone Receptor Gene. Arch. Gen. Psychiatry 2008, 65 (2), 190-200. DOI: 10.1001/archgenpsychiatry.2007.26
44. Grabe, H. J.; Schwahn, C.; Appel, K.; Mahler, J.; Schulz, A.; Spitzer, C.; Fenske, K.; Barnow, S.; Lucht, M.; Freyberger, H. J.; John, U.; Teumer, A.; Wallaschofski, H.; Nauck, M.; Völzke, H. Childhood Maltreatment, the Corticotropin-Releasing Hormone Receptor Gene and Adult Depression in the General Population. Am. J. Med. Genet. B Neuropsychiatr. Genet. 2010, 1538 (8), 1483-1493. DOI: 10.1002/ajmg.b.31131
45. Kranzler, H. R.; Feinn, R.; Nelson, E. C.; Covault, J.; Anton, R. F.; Farrer, L.; Gelernter, J. A CRHR1 Haplotype Moderates the Effect of Adverse Childhood Experiences on Lifetime Risk of Major Depressive Episode in African-American Women. Am. J. Med. Genet. B Neuropsychiatr. Genet. 2011, 1568 (8), 960-968. DOI: 10.1002/ajmg.b.31243
46. Polanczyk, G.: Caspi, A.; Williams, B.; Price, T. S.; Danese, A.; Sugden, K.; Uher, R.; Poulton, R.; Moffitt, T. E. Protective Effect of CRHR1 Gene Variants on the Development of Adult Depression Following Childhood Maltreatment: Replication and Extension. Arch. Gen. Psychiatry 2009, 66 (9), 978-985. DOI: 10.1001/archgenpsychiatry.2009.114
47. Heim, C.; Bradley, B; Mletzko, T. C.; Deveau, T. C; Musselman, D. L; Nemeroff, C. B.; Ressler, K. J.; Binder, E. B. Effect of Childhood Trauma on Adult Depression and Neuroendocrine Function: Sex-Specific Moderation by CRH Receptor 1 Gene. Front. Behav. Neurosci. 2009, 3, 41. DOI: 10.3389/neuro.08.041.2009
48. Amstadter, A. B.; Nugent, N. R.; Yang, B.-Z.; Miller, A.; Siburian, R.; Moorjani, P: Haddad, S; Basu, A.; Fagerness, J.; Saxe, G.; Smoller, J. W.; Koenen, K. C. Corticotrophin-Releasing Hormone Type 1 Receptor Gene (CRHR1) Variants Predict Posttraumatic Stress Disorder Onset and Course in Pediatric Injury Patients. Dis. Markers 2011, 30 (2-3), 89-99. DOI: 10.3233/DMA-2011-0761
49. White, S.; Acierno, R.; Ruggiero, K. J.; Koenen, K. C.; Kilpatrick, D. G.; Galea, S.; Gelernter, J; Williamson, V.; McMichael, O: Vladimirov, V. I.; Amstadter, A. B. Association of CRHR1 Variants and Posttraumatic Stress Symptoms in Hurricane Exposed Adults. J. Anxiety Disord. 2013, 27 (7). 678-683. DOI: 10.1016/jandis.2013.08.003
50. You, H.; Lu, B. Diverse Functions of Multiple BDNF Transcripts Driven by Distinct BDNF Promoters. Biomolecules 2023, 13(4), 655. DOI: 10.3390/biom13040655
51. Zhang, L.; Li, X-X: Hu, X.-Z. Post-traumatic Stress Disorder Risk and Brain-Derived Neurotrophic Factor Val66Met. World J. Psychiatry 2016, 6(1), 1-6. DOI: 10.5498/wjp.v6.i1.1
52. Felmingham, K. L.: Dobson-Stone, C.; Schofield, P. R.; Quirk, G. J.; Bryant, R. A. The Brain-Derived Neurotrophic Factor Val66Met Polymorphism Predicts Response to Exposure Therapy in Posttraumatic Stress Disorder. Biol. Psychiatry 2013, 73 (11), 1059–1063. DOI: 10.1016/ j.biopsych.2012.10.033
53. Jin, M. J.; Jeon, H.; Hyun, M. H.; Lee, S.-H. Influence of Childhood Trauma and Brain-Derived Neurotrophic Factor Val66Met Polymorphism on Posttraumatic Stress Symptoms, and Cortical Thickness. Sci. Rep. 2019, 9(1), 6028. DOI: 10.1038/s41598-019-42563-6

DOI: 10.36838/v7i10.138

54. Correia, A. S.: Cardoso. A.: Vale, N. BDNF Unveiled: Exploring Its Role in Major Depression Disorder Serotonergic Imbalance and Associated Stress Conditions. Pharmaceutics 2023, 15 (8), 2081. DOI: 10.3390/pharmaceutics15082081
55. NHS. Talking Therapies and Counselling; National Health Service: U.K., published January 25, 2021. https://www.nhs.uk/mental-health/talking- therapies-medicine-treatments/talking-therapies-and-counselling/
56. NHS, Overview - SSRI Antidepressants; NHS published 2021. https://www.nhs.uk/mental-health/talking-therapies-medicine-treatments/ medicines-and-psychiatry/ssri-antidepressants/overview/
57. Schrader, C.; Ross, A. A Review of PTSD and Current Treatment Strategies. Mo. Med. 2021, 118 (6), 546-551.
58. Skolariki, K.; Vlamos, P. Exploring Gene-Drug Interactions for Personalized Treatment of Post-traumatic Stress Disorder, ProQuest, 2024. DOI: 10.3389/fncom.2023.1307523
59. Mosiołek, A.; Mosiołek, J.; Jakima, S.; Pięta, A.; Szulc, A. Effects of Antidepressant Treatment on Neurotrophic Factors (BDNF and IGF-1) in Patients with Major Depressive Disorder (MDD). J. Clin. Med. 2021, 10 (15), 3377. DOI: 10.3390/jcm10153377
60. SLC6A3 Dep. Map Gene Summary. Depmap.org, published 2021. https://depmap.org/portal/gene/SLC6A3?tab=overview (accessed June 11, 2025)
61. Genesight. Get to Know a Gene: SLC6A4. GeneSight published on July 13, 2017. https://genesight.com/white-papers/get-to-know-a-gene-slc6a4/
62. Zhang, J.-P.: Lencz, T.; Malhotra, A. K. D2 Receptor Genetic Variation and Clinical Response to Antipsychotic Drug Treatment: A Meta-analysis:. Am. J. Psychiatry 2010, 167 (7), 763-772. DOI: 10.1176/appi.ajp.2009.09040598
63. Fawver, J.; Flanagan, M.; Smith, T.; Drouin, M.; Mirro, M. The Association of COMT Genotype with Buproprion Treatment Response in the Treatment of Major Depressive Disorder. Brain Behav. 2020, 10(7), e01692. DOI: 10.1002/brb3.1692
64. Huang, E.: Zai, C. C.; Lisoway, A.; Maciukiewicz, M.; Felsky, D.: Tiwari, A. K.; Bishop, J. R.; Ikeda, M.; Molero, P.; Ortuno, F.; Porcelli, S.: Samochowiec, J.; Mierzejewski, P: Gao, S.; Crespo-Facorro, B.; Pelayo-Terán, J. M.; Kaur, H.; Kukreti, R.; Meltzer, H. Y.; Lieberman, J. A.: Potkin, S. G.: Müller, D. J.; Kennedy, J. L. Catechol-O-Methyltransferase Val158Met Polymorphism and Clinical Response to Antipsychotic Treatment in Schizophrenia and Schizo-Affective Disorder Patients: A Meta-analysis. Int. J. Neuropsychopharmacol. 2016, 19 (5), pyv132. DOI: 10.1093/in/yv132
65. Ramoz, N.: Hoertel, N.; Nobile, B.; Voegeli, G.; Nasr. A.; Le Strat, Y.; Courtet, P.; Gorwood, P. Corticotropin Releasing Hormone Receptor CRHR1 Gene Is Associated with Tianeptine Antidepressant Response in a Large Sample of Outpatients from Real-Life Settings. Transl. Psychiatry 2020, 10 (1), 378. DOI: 10.1038/s41398-020-01067-y
66. Tafet, G. E.; Nemeroff, C. B. Pharmacological Treatment of Anxiety Disorders: The Role of the HPA Axis. Front. Psychiatry 2020, 11, 443. DOI: 10.3389/fpsyt 2020.00443
67. Zhang, G.-F.; Ren, Y.-P.; Sheng, L.-X.; Chi, Y.; Du, W.-J.; Guo, S.; Jiang, Z.-N.: Xiao, L.; Luo, X.-N.; Tang, Y.-L.; Smith, A. K.; Liu, Z.-Q.; Zhang, H.-X. Dysfunction of the Hypothalamic-Pituitary-Adrenal Axis in Opioid Dependent Subjects: Effects of Acute and Protracted Abstinence. Am. J. Drug Alcohol Abuse 2008, 34 (6), 760-768. DOI: 10.1080/00952990802385781
68. Nikolić, T.; Bogosavljević, M. V.: Stojković, T.; Kanazir, S.; Lončarević-Vasiljković, N.; Radonjić, N. V.: Popić, J.: Petronijević, N. Effects of Antipsychotics on the Hypothalamus-Pituitary-Adrenal Axis in a Phencyclidine Animal Model of Schizophrenia. Cells 2024, 13 (17), 1425. DOI: 10.3390/ cells13171425
69. Morland, L.; Woolley, J. Psychedelic-Assisted Therapy for PTSD, published August 27, 2024. http://www.ptsd.va.gov.https://www.ptsd.va.gov/ professional/treat/txessentials/psychedelics_assisted_therapy.asp
70. Study Indicates Stellate Ganglion Block Can Be Effective Treatment for PTSD: RTI, published November 1, 2019. https://www.rti.org/impact/study- indicates-stellate-ganglion-block-can-be-effective-treatment-ptsd
71. Steckler, T.; Risbrough, V. Pharmacological Treatment of PTSD - Established and New Approaches. Neuropharmacology 2012, 62 (2), 617-627. DOI: 10.1016/j.neuropharm.2011.06.012

Author

Sophia Tang is a senior at Boston University Academy, Boston, MA. She is interested in medicine and psychology. After taking a psychopathology course, this led to a combined interest in the human brain and mental health. She hopes to expand the depth of her knowledge in the future.

IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Corporate Greenwashing and Consumer Purchase Frequency in China Across Three Industries

Yufang He

8 Hao Jiantashan Lu Huangpu Qu Guangzhou Shi, Guangdong China; 1853647813@qq.com
Mentor: Zitong Li

ABSTRACT: Corporate greenwashing practices will jeopardize sustainable development efforts and deter consumer trust in the long run. Especially in China, there is a lack of policies regulating the occurrence of greenwashing, which refers to exaggerating, misleading, trimming, or presenting false information on the level of sustainability of a particular product, service, or corporate action. This paper identifies four prevalent forms of greenwashing: Good Cover Bad, Vague Wording, Dishonesty, and Irrelevancy. Researchers collected data from a self-designed survey asking consumers to rank their purchase frequency of a particular product from three industries, beauty products, food and beverage, and clothing, before and after the brand commits greenwashing in a hypothetical situation. A cluster analysis compares and contrasts their responses, clustering all 481 respondents into four groups based on their shared behavioral preferences. By cross-comparing the four groups, we make connections on how individual socio-economic background can influence their sensitivity to corporate greenwashing.

KEYWORDS: Environmental Engineering, Pollution Control, Policy, Corporate Greenwashing, Consumer Response.

Introduction

Greenwashing refers to exaggerating, misleading, trimming, or presenting false information on the level of sustainability of a particular product, service, or corporate action. It has become a more pressing and prevalent issue in China as societies grow in environmental awareness. Consumers' expectations for environmentally responsible products incentivized corporations to take the shortcut and greenwash instead of properly investing in sustainable production. Greenwashing undermines environmental actions; in particular, it is incompatible with the UN Sustainable Development Goal 13, "Climate Action," and the High-Level Expert Group's pledge to 'zero tolerance for net-zero greenwashing.'[1] If left unaddressed, greenwashing can impede environmental progress and erode consumer trust.[2] It poses a significant challenge to the larger economy and individuals living under the system. Companies that benefit from greenwashing tactics are less motivated to invest in truly sustainable practices.[3] Consumers who identify instances of greenwashing may overlook the fundamental principles of sustainable development and thus underestimate the need for sustainable goals. In China, no specialized laws have been enacted against corporate greenwashing. The very concept of "greenwashing" is still in its infancy stages, unaware to the greater public. Currently, China regulates greenwashing primarily through advertising and consumer protection laws.[4] Yet, these laws do not cover all the various forms of greenwashing, leaving many firms with opportunities to get away unpenalized.

Previously identified research gaps include a lack of quantitative analysis of consumer sensitivity to different forms of greenwashing and how individual socioeconomic profiles may influence that sensitivity; a shortage in the examination of other stakeholders' involvement in greenwashing aside from consumers and corporations; and an absence of a defining taxonomy that explores the difference in perceived severity of the various forms of greenwashing.[3] Furthermore, few pieces of literature have addressed consumers' responses to greenwashing, which is a critical force that incentivizes corporate activity. We identified consumer awareness as the most important breakthrough point in designing an effective policy. Corporations will stop greenwashing if their profits are declining as a result. Thus, this paper seeks to align definitions of greenwashing by categorizing four prevalent greenwashing practices and localizing the taxonomy under the Chinese economy. Also, this paper aims to fill research gaps by conducting an in-depth analysis of greenwashing in three industries: beauty products, clothing, food, and beverage, all of which are frequent, daily products, and examines the difference in the impact of four identified forms of greenwashing on consumer purchase frequency. Last, concluding from the survey results, this paper will suggest implications for stakeholders to understand, identify, preempt, and tackle greenwashing across industries.

Literature Review:

As societies become more environmentally aware, more protocols, such as the Corporate Social Responsibility (CSR) and the ESG (environmental, social, governance) criteria, have emerged as new evaluations for business investments. Data show that many consumers expect companies to demonstrate high social involvement, be environmentally active, and operate with high ethical conduct.[5] To work up to that expectation, many firms, taking advantage of the underdeveloped environmental legal system, adapt to greenwashing practices where they exaggerate the extent of their sustainability level or pres-

DOI: 10.36838/v7i10.148

ent misinformation to attract green consumers and create a more favorable brand image.[5]

Definition of Greenwashing:

Corporate greenwashing is the act of exaggerating, misleading, trimming, or presenting false information on the level of sustainability of a particular product, service, or corporate action.[3] When discovered, greenwashing affects corporations in several ways. It lowers investors' intention to invest and damages consumers' trust, consequently lowering purchase intention. For other competing firms, greenwashing by one corporate can ruin the reputation of the entire industry, which H. Wang *et al.* referred to as the spillover effect.[6]

Four main categories of greenwashing are identified in this paper:

Vague Wording. Vague wording refers to when firms tag their product/service with environmentally friendly labels yet do not provide authorized documentation to back up their claim. For example, in 2022, H&M was sued in New York for featuring an "environmental scoreboard" showing a clothing section produced with sustainable material. However, critics point out that the "scoreboard" characterizes certain items as being more sustainable than they are.[7]

Good covering, bad. Good cover, bad is a form of corporate greenwashing when firms intentionally display an incomplete record of their green credentials, showing only the favorable aspects of a product/service, or divert sustainability claims to cover a questionable environmental record. For example, HSBC, one of the largest consumer banks, had been advertising its self-claimed environmental act of participating in a Net Zero alliance while intentionally avoiding mentioning their finances in a fossil fuel project. For this, HSBC had been officially called out by the Advertising Standards Authority (ASA) for committing greenwashing.[8]

Dishonesty. Dishonesty, as a form of greenwashing, exists when firms provide false or made-up data as any form of qualification for their sustainability or when an inconsistency exists between corporate actions and stated intentions. For example, Coca-Cola once claimed its bottle packaging was 100% recyclable. Still, an environmental organization, the Sierra Club, points out that the sticker on the outside of the bottle is made of polypropylene, a non-recyclable material.[9]

Irrelevance. Companies make technically compelling but environmentally irrelevant claims.

Sectoral Impact of Greenwashing:

This research considers three consumer-related industries: food, clothing, and beauty. These three industries are chosen for their regular association with consumer choice, and they account for a significant proportion of overall consumer spending.[10, 11] These three industries are chosen for reasons as followed.

Greenwashing practiced by corporations in the food industry can be highly implicit. Colors and images are less detectable elements, yet they convey hidden and significant meanings that can influence a consumer's purchase decision.[12, 13] For example, the color green on product packaging is often associated with concepts of nature and safety. Consumers tend to perceive it as a cue signaling an environmentally friendly product or brand, even when the product itself might not qualify as being sustainable.[14] In particular, research conducted by Boncinelli *et al.*[15] proves that while green packaging, the practice of using visually stimulating elements such as colors and images on food packages, can mislead consumer perception of the level of sustainability of a product, it can potentially increase the market share of a firm, incentivizing greenwashing. The involvement of multiple academic expertise, including psychology, biology, and behavioral economics, poses a noteworthy challenge to policymakers when combating greenwashing in the food industry. For such reason, this research aims to uncover the implicit factors influencing consumers' sensitivity toward greenwashing on food products, and how policymakers could take advantage of this knowledge to better inform the public and discourage consumption.

The second industry examined in this study is the clothing industry, which was chosen for its high susceptibility to greenwashing. According to the EU, textile production doubled from 2000 to 2015, and its consumption is expected to increase by 63% by 2030.[16] This industry, one of the most environmentally unfriendly industries, must adapt to a more sustainable production process.[16] Currently, the EU has proposed several strategies to regulate the sustainability of the textile industry. First, the release of polluted water in manufacturing processes was banned. Then, a reduction in microplastics to alleviate the impact of microfiber shedding was required, and the standardization of green reports further enhanced the validity of sustainable development in the fashion industry.[16]

The social and policy expectations put on the fashion industry to be more sustainable can backfire and lead to more greenwashing.[15] Greenwashing actions in this industry include overstating the level of sustainability of a firm or product and invalid eco-labeling, which, when unidentified, can help improve brand image and incentivize guilt-free consumption. Thus, this study sought to investigate more effective policies to hinder greenwashing by understanding the consumer decision-making process in the clothing industry.

Finally, the cosmetics industry was also chosen for its high susceptibility to greenwash. Research done by Rocca *et al.*[17] claims that cosmetics, considering the enormous participation in the global resource and consumer market, have only recently entered the sustainability market. Information-wise, we are witnessing an increase in the number of articles published on sustainability in the cosmetics industry from 2010 onwards. More consumers are aware of the issue and start looking for sustainable products. Corporations adapting to sustainable marketing strategies are guaranteed an advantage over their competitors and obtain benefits in terms of brand reputation and customer satisfaction.[18] However, this lures businesses to promote sustainability without actually investing in sustainable production, leading to greenwashing in the cosmetics industry. Corporations consider sustainability to be a social expectation enforced by stakeholders or the government rather than a corporate initiative.[18]

It is important that policymakers have a better understanding of what drives consumers to respond to greenwashing practices in the cosmetics industry in order to more effectively respond to growing malpractices, which are driven by demands for more sustainable products.

Consumer Response to Greenwashing:

In general, corporate greenwashing has the following effects. Since most manifestations of greenwashing appear in product/business advertisements, greenwashing lowers consumers' willingness to purchase the associated product, and they are less likely to trust advertisements accordingly. Greenwashing deters brand credibility; once lost, it becomes challenging to reestablish brand loyalty and consumer brand engagement. An interesting effect of corporate greenwashing regards its spillover effect, referring to when one brand greenwashes, consumers are less likely to buy similar products in the same industry, lowering the overall judgment of the industry.[19]

Several methods are proposed, and some have been implemented to cope with greenwashing. First, the emergence of environmental organizations such as Greenpeace's StopGreenwash.org serves to regulate the CSR reports. Second, promoting education for sustainable development ensures consumers can detect misleading information. Third, examining CSR reports by a credible third party assists in providing valid documentation for corporations' claimed sustainability.[5]

■ Methods

To meet the stated research objectives, this paper conducts a literature review on existing research on greenwashing. Conclusions drawn from the literature review are used to place the topic of greenwashing in a broader economic background and refine arguments made in this paper. Additionally, this paper will both quantitatively and qualitatively analyze data collected from a self-designed survey conducted in August 2023. We spread the survey on social media, opened it up to all Chinese internet users, and received a total of 481 valid responses (n=481) from these online consumers in China. The survey is structured into three sections: consumer socioeconomics profile, including age, gender, income, occupation, status, and awareness of greenwashing prior to completing the survey; consumer purchase frequency by product type, including luxury and necessity products from clothing, beauty products, and food and beverage industry, without corporate greenwashing; and purchase frequency of the same products under a hypothetical situation when brands commit each of the four, identified forms of greenwashing. The response in the third section serves as a measurement of consumer sensitivity toward corporate greenwashing. Considering the difference in socioeconomic background of individual respondents, this survey is designed to answer the research question: Which factors most significantly influence consumer purchase decisions in response to corporate greenwashing? And how that may vary depending on respondents' different socioeconomic backgrounds. This paper is separated into two sections.

This paper will dissect a cluster analysis that categorized the respondents into four groups based on similarities in their socio-economic profile and their responses to different forms of corporate greenwashing. The analysis was built based on first-hand data collected in the survey. These four groups identify behavioral preferences in consumers with distinct socioeconomic characteristics and provide insights into policy implications. In simple conclusion, three sets of matrixes are being compared in this paper, each contributing to the final cluster analysis: first, consumer purchase frequency of a particular product before and after a corporation commits greenwashing; second, cross-comparison between the four forms of greenwashing, and how consumers sensitivity alters from one to another; and third, cross-comparison of products from three different industries, and how consumers respond differently depending on different product types.

Based on the survey results, a 5-point scale is designed to categorize consumer sensitivity specialized in the scope of this study.

- Not sensitive: 100% of respondents report no impact of greenwashing on consumer decision
- Modestly sensitive: over 75% of respondents report no impact or minor impact of greenwashing on consumer decision-making
- Moderately sensitive: over 50% of respondents report a significant impact of greenwashing on consumer decision-making
- Sensitive: over 75% of respondents report a significant impact of greenwashing on consumer decision-making
- Extremely sensitive: All respondents report a significant impact of greenwashing on consumer decision-making

■ Survey Results and Discussion

Economic analysis of Chinese consumers' perception of greenwashing

The survey has received a total of 481 valid samples (n=481), with the following results:

Behavior: Response to Greenwashing:

Purchase Frequency Without Perceived Greenwashing:

The first section of the survey asks respondents about their regular purchase frequency of beauty products, clothing products, and food & beverage products. Each product is divided into subcategories: necessity products (products consumers purchase regardless of income level) and luxury product (non-life essentials products purchased with excess income). The results are summarized into the following points:

1. Most respondents purchase necessary beauty products on a seasonal and monthly basis.
2. The purchase of luxury beauty products is slightly less frequent than that of necessary beauty products.
3. Most respondents purchase necessary clothing products on a seasonal and monthly basis.
4. The frequency of purchases of luxury clothing products is scattered, with no distinct trend.
5. Most respondents purchase necessary food and beverage products weekly and monthly.
6. The purchase frequency of luxury food and beverage products, such as wine, is evenly distributed and slightly less frequent than other products. More people are in the "no purchase" or "seasonal purchase" group.

DOI: 10.36838/v7i10.148

Cluster Analysis

The following cluster analysis has identified four groups of respondents with shared traits based on their purchase frequency response to corporations committing each of the four types of greenwashing.

(1) Group 1: "Senior Eco-Enthusiasts":

Group 1 consists of slightly more females than males, with the majority aging from 26 to 40. 70.6% of the respondents currently work full-time, with 16.5% working in administrative positions and 16.2% in purchasing. Most respondents have three or more cohabitants and are modestly familiar with greenwashing, with a quarter of the respondents having already taken action against it. Out of the quarter who have taken action, 52% have taken action involving a cease in purchasing products from the greenwashing brand, and 45.6% have decided to share their environmental knowledge with family and friends. (*Cohabitant: includes family members living in the same household and dependents, aka people financially dependent on the respondent)

(2) Group 2 "Sustainable Professionals":

Group 2 consists of 20% more females than males, with the majority aging from 18 to 25 and 31 to 50. 61.2% of the group currently works full-time, with 17.3% occupying administrative positions and 15.3% in marketing. 67.3% of the respondents have three or more cohabitants, and slightly over half of the group is unaware of greenwashing, with more than a quarter of the population being entirely unfamiliar with the concept. Out of the aware half, 37.8% have taken action advocating against greenwashing through a cease in purchasing products from the greenwashing brand.

(3) Group 3 "Green Beginners":

Group 3 consists of an equal percentage of males and females and a considerable 5.3% percentage of non-binary gender. The age group is polarized with the majority aging from 18-25 and 41-50, respectively. More than half of the respondents currently work full-time, and 36.8% are students. 57.9% of the respondents have three or more cohabitants. Slightly over half of the respondents in Group 3 are modestly aware of greenwashing. Compared to the previous two groups, this group has a lower percentage of respondents who had taken action against greenwashing; nevertheless, most action-takers avoided purchasing greenwashing brands.

(4) Group 4 "Selectively Active Elites":

Respondents in this group are distinctively characterized. Group 4 has more females (66.7%) than males (33.3%). All respondents fall into the 31- to 50-year-old age cohort and work full-time in either a marketing, administrative, or teaching position. This group comprises the well-off, knowledgeable individuals, with the majority having three or more cohabitants. Surprisingly, 66.6% of the group is unfamiliar with greenwashing, yet the remaining 33.3% is highly active against greenwashing practices. Group 4 has taken action, including educating their families and friends about greenwashing and directly contacting the responsible brand, which speaks to the social influence and resources this group has at hand.

Greenwashing Analysis: consumer responses to different industries:

The survey breaks down group-based consumer responses to the three industries on a 5-point scale.

Insensitive: Less than 25% of respondents report a significant impact of greenwashing on consumer decisions.

Modestly sensitive: Over 25% of respondents report a significant impact of greenwashing on consumer decisions.

Moderately sensitive: Over 50% of respondents report a significant impact of greenwashing on consumer decisions.

Sensitive: over 75% of respondents report a significant impact of greenwashing on consumer decisions.

Extremely sensitive: All respondents report a significant impact of greenwashing on consumer decisions.

Table 1.0.: Group 1's reaction to Greenwashing in the Beauty Industry. Group 1 respondents are highly sensitive toward all four forms of corporate greenwashing practices in the beauty industry.

(Beauty Product Industry) Forms of Greenwashing	No effect (%)	Minor impact on brand perception but will continue purchase (%)	Significant effect, seeking for substitutes (%)	Significant effect: avoid purchasing from this brand (%)	Significant effect will switch brand immediately (%)
Vague Wording	0.9	4.3	21.1	48.9	24.8
Good Cover Bad	0.0	4.3	23.5	38.8	33.3
Dishonesty	0.9	2.4	23.5	38.5	34.6
Irrelevancy	0.9	6.1	24.8	43.4	24.8

Table 1.1.: Group 2's reaction to Greenwashing in the Beauty Industry. Group 2 respondents are less sensitive to corporate greenwashing practices in the beauty industry than Group 1. There are significantly fewer consumers who will respond to greenwashing by immediately switching brands.

(Beauty Product Industry) Forms of Greenwashing	No effect (%)	Minor impact on brand perception but will continue purchase (%)	Significant effect, seeking for substitutes (%)	Significant effect: avoid purchasing from this brand (%)	Significant effect will switch brand immediately (%)
Vague Wording	6.1	32.7	48.0	11.2	2.0
Good Cover Bad	1.0	19.4	65.3	14.3	0.0
Dishonesty	1.0	12.2	45.9	29.6	11.2
Irrelevancy	7.1	24.5	58.2	8.2	2.0

a. Analysis:

A comparison between Table 3.0 and Table 3.1 shows that Group 1 is more sensitive to all forms of greenwashing in the beauty product industry than Group 2. This trend could be derived from the difference in gender composition in the two groups. Group 2 consists of more females who purchase beauty products more frequently and may continue to purchase favorable and established brands even if they commit greenwashing. Group 2 respondents have a more inelastic demand relative to greenwashing for beauty products.

A pattern exists in both groups where Irrelevancy appears to be the most tolerated form of greenwashing. This implies that consumers perceive Irrelevancy as a less severe act of greenwashing than the other three forms. Additionally, Dishonesty is the most sensitive form of greenwashing for both groups. Dishonesty is perceived as the most severe and unforgivable act, since it's a deliberate act of deception by companies. Con-

sumers value integrity when deciding whether they want to purchase from a certain brand.

Table 1.2.: Group 1's Reaction to Greenwashing in the Food and Beverage Industry (FB). Group 1 respondents are highly sensitive toward all four forms of corporate greenwashing practices in the FB industry, similar to their response to beauty products.

(Food and Beverage Industry) Forms of Greenwashing	No effect (%)	Minor effect on brand perception but will continue purchase (%)	Significant effect, seeking for substitutes (%)	Significant effect: avoid purchasing from this brand (%)	Significant effect will switch brand immediately (%)
Vague Wording	0.9	5.2	21.4	39.1	33.3
Good Cover Bad	0.0	4.0	21.4	42.2	32.4
Dishonesty	0.0	2.8	20.8	41.3	35.2
Irrelevancy	0.0	4.0	26.6	38.8	30.6

Table 1.3.: Group 2's Reaction to Greenwashing in the Food and Beverage Industry. Group 2 respondents display a lower sensitivity toward corporate greenwashing practices in the FB industry compared to group 1. There are significantly fewer consumers who will respond by immediately switching brands.

(Food and Beverage Industry) Forms of Greenwashing	No effect (%)	Minor impact on brand perception but will continue purchase (%)	Significant effect, seeking for substitutes (%)	Significant effect: avoid purchasing from this brand (%)	Significant effect will switch brand immediately (%)
Vague Wording	4.1	27.6	48.0	16.3	4.1
Good Cover Bad	0.0	15.3	55.1	25.5	4.1
Dishonesty	5.1	22.4	52.0	12.2	8.2
Irrelevancy	7.1	20.4	54.1	18.4	0.0

a. Analysis:

In general, Group 2 is more sensitive toward greenwashing behaviors in the Food and Beverage industry than in the Beauty Products industry (Table 3.3). This could be explained by considering the majority female presence in Group 2, of which the majority has a family make-up of three or more cohabitants. The fact that females tend to take responsibility for most housekeeping chores, like food and beverage purchasing, could explain why they are more sensitive toward products they purchase for themselves and their families on a frequent basis. Potentially, people may associate product quality with greenwashing and refuse to buy a greenwashed product, perceiving it as of worse quality, especially when it deals with food and beverages.

Good Cover Bad is an obvious violation of consumer trust. It is also one of the most sensitively responded to forms of greenwashing, demonstrating the significance of consumer trust in purchase decisions, as shown in Tables 3.2 and 3.3.

Table 1.4.: Group 1's Reaction to Greenwashing in the Clothing Industry (CLO). Group 1 respondents are highly sensitive toward all four forms of corporate greenwashing practices in the CLO industry, similar to their response to beauty products and FB products.

(Clothing Industry) Forms of Greenwashing	No effect (%)	Minor impact on brand perception but will continue purchase (%)	Significant effect, seeking for substitutes (%)	Significant effect: avoid purchasing from this brand (%)	Significant effect will switch brand immediately (%)
Vague Wording	0.6	6.4	20.8	42.8	29.4
Good Cover Bad	0.0	5.5	17.4	45.3	31.8
Dishonesty	0.0	4.3	26.0	39.1	30.6
Irrelevancy	0.6	4.6	22.9	44.3	27.5

Table 1.5.: Group 2's Reaction to Greenwashing in the Clothing Industry. Group 2 respondents are highly sensitive toward all four forms of corporate greenwashing practices in the CLO industry, different from their response to beauty products and FB products.

(Clothing Industry) Forms of Greenwashing	No effect (%)	Minor impact on brand perception but will continue purchase (%)	Significant effect, seeking for substitutes (%)	Significant effect: avoid purchasing from this brand (%)	Significant effect will switch brand immediately (%)
Vague Wording	0.9	4.3	21.1	48.9	24.8
Good Cover Bad	0	4.3	23.5	38.8	33.3
Dishonesty	0.9	2.4	23.5	38.5	34.6
Irrelevancy	0.9	6.1	24.8	43.4	24.8

a. Analysis:

Group 1 is more sensitive to the Good Cover Bad form of greenwashing, as shown in Table 4.4, and Group 2 is more sensitive to the Dishonesty form of greenwashing, as shown in Table 4.5. This appears to be the opposite of the tendency of consumer reaction towards greenwashing in the Food and Beverage industry. This indicates that consumer response to the same form of greenwashing may change depending on the product.

Table 1.6.: Group 3's Reaction to Greenwashing in the Beauty Product Industry. Group 3 respondents are insensitive to the practice of vague wording as a form of corporate greenwashing in the beauty industry and modestly sensitive toward the other three forms.

(Beauty Product Industry) Forms of Greenwashing	No effect (%)	Minor impact on brand perception but will continue purchase (%)	Significant effect, seeking for substitutes (%)	Significant effect: avoid purchasing from this brand (%)	Significant effect will switch brand immediately (%)
Vague Wording	36.8	54.4	8.8	0.0	0.0
Good Cover Bad	29.8	31.6	35.1	3.5	0.0
Dishonesty	29.8	35.1	22.8	8.8	3.5
Irrelevancy	26.3	33.3	31.6	3.5	5.3

Table 1.7.: Group 4's Reaction to Greenwashing in the Beauty Product Industry. Group 4 consumers are completely insensitive to all four forms of corporate greenwashing in the beauty industry.

(Beauty Product Industry) Forms of Greenwashing	No effect (%)	Minor effect on brand perception but will continue purchase (%)	Significant effect: seeking for substitutes (%)	Significant effect: avoid purchasing from this brand (%)	The significant impact will switch brands immediately (%)
Vague Wording	100.0	0.0	0.0	0.0	0.0
Good Cover Bad	100.0	0.0	0.0	0.0	0.0
Dishonesty	100.0	0.0	0.0	0.0	0.0
Irrelevancy	100.0	0.0	0.0	0.0	0.0

a. Analysis:

Due to their limited awareness of sustainability and green consumption, Group 3 individuals are reluctant to act against any form of greenwashing in the beauty product industry (Table 3.6). They simply cannot distinguish greenwashed products from truly sustainable products and are unaware of the consequences of such malpractice.

Group 4's distinct behaviors are characterized by their complete negligence of greenwashing in the beauty industry, demonstrating their perfectly inelastic demand for such products (Table 4.7). Considering the female-dominating composition of Group 4, many of whom also work as corporate administrators, it becomes clear that these respondents have a strong preference for their chosen brands of beauty products

DOI: 10.36838/v7i10.148

and that environmental concerns are neither compelling nor important enough to alter their purchasing behavior.

Table 1.8.: Group 3's Reaction to Greenwashing in the Food and Beverage Industry. Group 3 respondents are insensitive to the practice of vague wording and irrelevance as a form of corporate greenwashing in the FB industry and modestly sensitive toward the other three forms.

(Food and Beverage Industry) Forms of Greenwashing	No effect (%)	Minor impact on brand perception but will continue purchase (%)	Significant effect, seeking for substitutes (%)	Significant effect: avoid purchasing from this brand (%)	Significant effect will switch brand immediately (%)
Vague Wording	35.1	47.4	15.8	1.8	0.0
Good Cover Bad	26.3	42.1	15.8	14.0	1.8
Dishonesty	29.8	36.8	26.3	3.5	3.5
Irrelevancy	28.1	50.9	14.0	7.0	0.0

Table 1.9.: Group 4's Reaction to Greenwashing in the Food and Beverage Industry. Group 3 respondents are insensitive to the practice of good cover bad as a form of corporate greenwashing in the FB industry, and modestly sensitive toward the other three forms.

(Food and Beverage Industry) Forms of Greenwashing	No effect (%)	Minor impact on brand perception but will continue purchase (%)	Significant effect, seeking for substitutes (%)	Significant effect: avoid purchasing from this brand (%)	Significant effect will switch brand immediately (%)
Vague Wording	33.3	0.0	0.0	0.0	66.7
Good Cover Bad	33.3	66.7	0.0	0.0	0.0
Dishonesty	33.3	0.0	0.0	33.3	33.3
Irrelevancy	33.3	0.0	33.3	0.0	33.3

c. Analysis:

Group 4 is significantly more sensitive and active against corporate greenwashing in the Food and Beverage industry, as shown in Table 3.9, than in the Beauty Product Industry. This could be explained considering the Group 4 respondents' pursuit of a healthy lifestyle and highly disciplined diet. (Assumption made based on their economic make-up) Their demand for distinctly high-quality food and beverages accounts for their rejection of any form of greenwashing in food and beverages. (In other words, they are financially capable and mentally aware of their diet.) This also hints at how consumers may associate greenwashing with product quality.

There is no notable tendency regarding Group 3's reaction to greenwashing in the Food and Beverage industry (Table 4.8). They are neither particularly active nor responsive, though their actions match the previously made assumption that they are "green beginners" who have just been introduced to the concept of greenwashing.

Table 1.10.: Group 3's Reaction to Greenwashing in the Clothing Industry. Group 3 respondents are insensitive to vague wording and irrelevance as a form of corporate greenwashing in the clothing industry, and modestly sensitive toward the other three forms.

(Clothing Industry) Forms of Greenwashing	No effect (%)	Minor impact on brand perception but will continue purchase (%)	Significant effect: seeking for substitutes (%)	Significant effect: avoid purchasing from this brand (%)	Significant effect will switch brand immediately (%)
Vague Wording	42.1	49.1	8.8	0.0	0.0
Good Cover Bad	31.6	38.6	21.1	8.8	0.0
Dishonesty	35.1	36.8	28.1	0.0	0.0
Irrelevancy	35.1	45.6	15.8	3.5	0.0

Table 1.11.: Group 4's Reaction to Greenwashing in the Clothing Industry. Group 4 consumers are completely insensitive to all four forms of corporate greenwashing in the clothing industry.

(Clothing Industry) Forms of Greenwashing	No effect (%)	Minor impact on brand perception but will continue purchase (%)	Significant effect: seeking for substitutes (%)	Significant effect: avoid purchasing from this brand (%)	Significant effect will switch brands immediately (%)
Vague Wording	0.0	0.0	33.3	0.0	66.7
Good Cover Bad	0.0	0.0	0.0	33.3	66.7
Dishonesty	0.0	0.0	0.0	0.0	100.0
Irrelevancy	0.0	0.0	0.0	66.7	33.3

c. Analysis:

Group 4 is significantly more sensitive towards greenwashing in the Clothing Industry than Group 3, as shown in Table 3.10 and Table 3.11. It can thus be inferred that Group 4 respondents have a more elastic demand for clothing products. Building off that conclusion, it can be assumed that Group 4, with a major percentage of respondents in positions of power, has other priorities, such as monetary benefit, that shape their evaluation of greenwashing severity in different products. A notable tendency shows that Group 4 respondents are particularly sensitive to greenwashing when corporations manifest brand dishonesty, compared to the other three groups.

■ Conclusion

This research reveals the complex relationship between consumer purchase behaviors and corporate greenwashing across three industries in China. Our findings show that consumers can be clustered into four groups based on their sustainability awareness, socioeconomic status, and attitude toward greenwashing. Different forms of greenwashing can also be categorized into four groups: Dishonesty, Good Cover Bad, Irrelevancy, and Vague Wording. Dishonesty and Good Cover Bad elicit the strongest consumer repulsion during purchase for Group 1, Group 2, and Group 4 consumers, while Vague Wording and Irrelevance are more tolerated, especially among consumers with inelastic demand or limited awareness. Additionally, consumers from Group 4 appear to be completely insensitive toward all four forms of greenwashing in the beauty industry, possibly due to high adherence to established brand values. This study underscores the importance of targeted policy measures. One-size-fits-all regulations are unlikely to hinder greenwashing effectively. Instead, policies should incorporate educational outreach, third-party certification, and stronger legal qualifications tailored to consumer group dynamics and industry characteristics.

Limitations:

Several limitations exist in the research methods. First, the survey collected 481 valid responses, which may not fully represent China's diverse consumer population. Survey responses were based on hypothetical scenarios, which may not accurately capture consumers' real-life purchase behaviors. For a better understanding of consumer behaviors, future researchers could conduct offline, real-time experiments given the proper resources and financial support. Additionally, this study only examined three industries (beauty, food and beverage, and

clothing). While these sectors are consumption-heavy, other industries like electronics or automobiles may exhibit different greenwashing dynamics.

Nonetheless, results from this study suggest that policymakers to combat greenwashing by raising consumer awareness on the deceptive nature of greenwashing practices. When consumers realize and stop the purchase of greenwashed products, it will discourage companies from greenwashing more immediately than what many current policies are doing. Future research should focus on evaluating country and industry-specific policies and examining the role of different stakeholders in preventing and addressing corporate greenwashing. Additionally, further research into how distinct socioeconomic factors—such as income level, social status, and family composition—influence each consumer cluster could help policymakers design more targeted and effective regulations.

Acknowledgments

I want to thank my mentor, Ms. Zitong Li, for taking the time to guide me, step-by-step, from the survey building to academic essay writing and eventually producing this original paper. For further interest, please check out the original survey design for this research.

https://www.wjx.cn/vm/wxAIPNe.aspx#

References

1. Greenwashing – the deceptive tactics behind environmental claims. https://www.un.org/en/climatechange/science/climate-issues/greenwashing (accessed Nov 10, 2023), *United Nations*.
2. Ioannou, I.; Kassinis, G.; Papagiannakis, G. How greenwashing affects the bottom line. https://hbr.org/2022/07/how-greenwashing-affects-the-bottom-line (accessed Nov 10, 2024), *Harvard Business Review*.
3. Santos, C.; Coelho, A.; Marques, A. A Systematic Literature Review on Greenwashing and Its Relationship to Stakeholders: State of Art and Future Research Agenda. Manag. Rev. Q. [Online] **2023**,1-25. https://doi.org/10.1007/s11301-023-00337-5. (accessed Sep 25, 2023).
4. Sustainability claims and greenwashing in China. https://cms.law/en/cze/publication/cms-green-globe/china (accessed Jan 22, 2024).
5. Wolniak, R. Relations between Corporate Social Responsibility Reporting and the Concept of Greenwashing. *Zesz. Nauk. - Politech. Łódz.*, Organ. Zarz. [Online] **2016**, 443–453.
6. Ahmad W.; Zhang Q. Green purchase intention: Effects of electronic service quality and customer green psychology. *J. Clean. Prod.* [Online] **2020**, 267. https://doi.org/10.1016/j.jclepro.2020.122053. (accessed Sep 25, 2023).
7. Adegeest, D.-A. H&M "greenwashed" its sustainability scores, says investigation. https://fashionunited.com/news/fashion/h-m-greenwashed-its-sustainability-scores-says-investigation/2022070448454 (accessed Jan 22, 2024).
8. Hicks, R. Eco-Business. https://bankonourfuture.org/news/hsbc-climate-action-festival/ (accessed Oct 21, 2023).
9. Sierra Club Sues Coca-Cola and Others Over Recyclability Claims. https://www.sierraclub.org/environmental-law/sierra-club-sues-coca-cola-and-others-over-recyclability-claims. (accessed Jan 1, 2024).
10. Eurostat. How much are households spending on food? https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20201228-1 (accessed Dec 22, 2023).
11. Ma, Y. Statista. China: Urban per capita spending on food 2022. https://www.statista.com/statistics/278858/per-capita-expenditure-of-private-households-in-china-on-groceries/ (accessed Dec 31, 2023).
12. Grandi, B; Cardinali, M.G. Colours and price offers: how different price communications can affect sales and customers' perceptions. *SSRN*. [Online] **2022**, 68, 1-16. https://doi.org/10.2139/ssrn.4052192 (accessed Dec 30, 2023).
13. Grossman, R.; Wisenbilt, J. What we know about consumers' color choices. *J. Market. Pract. Appl. Market Sci.* [Online] **1999**, 5, 78–88. https://www.emerald.com/insight/content/doi/10.1108/EUM0000000004565/full/html (accessed Dec 30, 2023).
14. Labrecque, L.I; Milne, G.R. Exciting Red and Competent Blue: The Importance of Color in Marketing. *J. Acad. Market. Sci.* [Online] **2012**, 40, 711–727. https://doi.org/10.1007/ (accessed Dec 30, 2023).
15. Boncinelli, F; Gerini, F; Piracci, G; Bellia, R; Casini, L. Effect of executional greenwashing on the market share of food products: An empirical study on green-colored packaging. *J. Clean. Prod.* [Online] **2023**, 391, 1-10. https://www.sciencedirect.com/science/article/pii/S095965262300416X?ref=pdf_download&fr=RR-2&rr=821d1377f9fa0976#bib22 (accessed Nov 09, 2023).
16. Brussels, E. C. European Commission. EU Strategy for Sustainable and Circular Textiles. https://ec.europa.eu/environment/publications/textiles-strategy_en (accessed Jan 2, 2023).
17. Rocca, R; Taisch, M; Acerbi, F; & Fumagalli, L. Sustainability Paradigm in the Cosmetics Industry: State of the art. *Elsevier Sci.* [Online] **2022**, 3, 1-14. https://www.sciencedirect.com/science/article/pii/S2772912522000574?ref=cra_js_challenge&fr=RR-1 (accessed Nov 9, 2023)
18. Eccles, R.G; Ioannou, I; Serafeim, G. The Impact of Corporate Sustainability on Organizational Processes and Performance. *Harv. Bus. Sch.* [Online] **2014**, 11. https://doi.org/10.1287/mnsc.2014.1984 (accessed Jan 2, 2024)
19. Wang, H; Ma, B; Bai, R; The spillover effect of greenwashing behaviours: An experimental approach. *Emerald Publishing Limited* [Online] **2019**, *3*, 283-295. https://www.emerald.com/insight/content/doi/10.1108/MIP-01-2019-0006/full/html (accessed Jan 22, 2024).

Authors

Yufang He is a sophomore at BASIS International School of Guangzhou. She is intrigued by studies that approach environmental issues and policies from an interdisciplinary perspective, just like the topic explored in this paper.

Ms. Zitong Li graduated from the University of Toronto with a dual degree, Master of Public Policy and Asia Pacific Studies. She has 10 years of research experience at the United Nations Development Programme, think tanks, NGOs, and sustainability consultancies, focusing on ESG, sustainable development, and corporate social responsibility.

DOI: 10.36838/v7i10.148

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Fighting the Obesity Epidemic: A Comprehensive Literature Review on Sugar Taxation

Vineet Vishnubhotla

Northwood High School, 4555 Portola Parkway, Irvine, California, USA; vineetvishnubhotla@gmail.com
Mentor: Gursel Aliyev

ABSTRACT: Sugar-sweetened beverage taxation has been adopted in many countries due to the adverse health outcomes associated with consuming excess sugar. Sugar is directly related to obesity and type 2 diabetes. Treating these diseases can put stress on healthcare systems. This review analyzes the effectiveness of sugar taxes in reducing consumption, identifies the most impactful tax rates, and explores which socioeconomic groups are most affected, along with how firms respond to government-imposed taxes.

KEYWORDS: Behavioral and Social Sciences, Economics, Supply and Demand.

Introduction

A study done by the WHO[1] found that people who consume 1-2 sugary beverages per day have a 26 percent greater chance of diabetes.

Obesity rates have risen over the past decade, primarily due to heavy sugar consumption. According to the American Heart Association,[2] the average American consumes 17 teaspoons of added sugar daily, 2-3 times the recommended amount. This amounts to about 60 lbs of added sugar annually, or 85 grams of added sugar daily. The most common sources of added sugar are sugar-sweetened beverages, coffee, desserts, and sweets.

Implementing sugar taxes serves public health goals and offers economic benefits. Over time, governments can cut healthcare costs and redirect healthcare expenditures towards improved education, improving infrastructure, and helping low-income households.

This paper explores the economic and public health rationale for sugar taxes. It seeks to answer the following research question: How do sugar taxes impact consumption patterns and public health outcomes in the United States?

Discussion

Supply & Demand Concepts:

In economics, supply refers to the amount a seller is willing to produce, while demand refers to the amount consumers want at a given price. When supply increases, it usually means more sellers in a given market. As such, prices tend to drop when supply increases, as there are close substitutes (i.e., other sellers).

Demand has an inverse relationship with price. When prices go up, demand falls, and vice versa. However, specific products, or services for that matter, can have different needs.

Demand may not react swiftly to price changes for more exclusive or sought-after items.

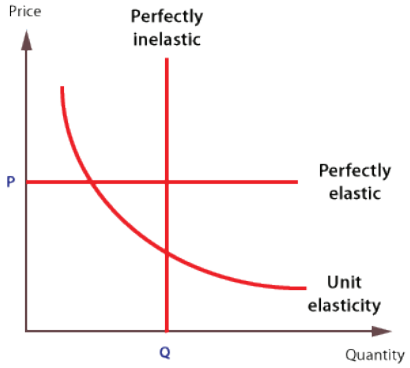


Figure 1: Price Elasticity Graph.

Price elasticities[3] are defined as the percent change in quantity demanded / price change. If demand were to change minimally for a large change in price, then there would be a small price elasticity of demand, or PE. Smaller PEs mean the demand for the goods is inelastic or less responsive to price. Larger PEs mean the demand for the goods is probably elastic or very responsive to price.

The absolute value of calculated PEs below 1 is inelastic, whereas anything above 1 would be considered elastic.

When price elasticity (PE) is less than 1 or inelastic, firms can transfer the tax burden by raising prices. However, when PE is less than 1, firms cannot effectively pass the tax onto consumers and must absorb it themselves, often leading to a revenue loss.

Price Elasticity of a Sugar-Sweetened Beverage Tax:

Recent studies in Mexico have found price elasticities ranging between -1.06 and -1.29 for soft drinks and -1.16 for SSBS (Sugar-Sweetened Beverages).[4]

Research[4] in Mexico used a difference-in-difference model to find that untaxed beverages showed a higher absolute difference, starting at 63m, following the immediate aftermath of sugar taxation. However, this positive disparity decreased over

DOI: 10.36838/v7i10.155

the 12-month tax period and became statistically insignificant by November 2014. Nonetheless, consumption still increased by 36 mils per capita per day. Sensitivity analysis may suggest a potential overestimation of 4 percent.

In Ecuador, study[5] found that price elasticity ranges between -1.17 and -1.33, depending on economic status.

In the UK,[6] price elasticity for sugar taxes ranges from 1.2 to 2.80. In response to government taxation, sellers have raised prices slightly, resulting in a notable 33.8% reduction in sugar consumption. This may indicate more elastic demand in the UK than in other countries worldwide.

Research done in Chile found a price elasticity of -1.37, corresponding to a 10% tax on sugary drinks.[7] This means that if the price increased by 10%, consumption would reduce by 13.7%.

A meta-analysis done by Escobar scanning the literature found a combined price elasticity estimate of -1.30.[8] Research also found that when prices rise, the consumption of fruit juices increases (0.388), whole milk increases (0.129), but there is a decrease in diet soft drinks (-0.423).

An analysis used regression models and controlled for demographic variables, income, and prices of other food and drink items to analyze the price elasticity of sugar taxation.[9] Brazil's adjusted price elasticity coefficient for household consumption of sugar-sweetened beverages (SSBs) was -0.85. This would mean that a 1.00% rise in SSB prices would correspond to a 0.85% decline in SSB consumption and signal inelastic demand.

While price elasticities measure consumer sensitivities, it is also important to evaluate consumption patterns post-tax implementation.

Post-Tax Sugar Beverage Consumption Rates:

To investigate the impact of Hungary's sugar tax, implemented in 2012, Pfinder used an Interrupted Time Series Analysis.[10] The study revealed a 4 percent reduction in the consumption of sugar-sweetened beverages subject to the tax. Although the research showed minimal risk of bias across various domains, the certainty of evidence for both primary and secondary outcomes was deemed very low. It's important to note that the tax did not encompass unprocessed sugar.

However, factors that can exist beyond taxes could contribute to declining consumption. In Ireland, for example, consumption decreased by 154 calories between 2009 and 2016, compared to a drop of 205 calories between 2003 and 2004, without taxation, highlighting consumers already switching to healthier choices.[11] The same journal also references a study using Nielsen Home Scanner Data to analyze sugar consumption rates.[12] However, this choice of methodology has several limitations, with one being that the data excludes consumption outside of the home.

In a comparison of two reviews, it was found that two studies, comprising a randomized controlled field experiment and a descriptive field study, targeted "less healthy" food and beverage categories in supermarkets. In a randomized controlled field experiment,[13] groups were assigned to either a control group or an experimental group, subjected to a 10% tax on "less healthy" foods and drinks, which included SSDS. Both studies found that while there was a short-term decrease in SSD purchases after one month of taxation, this effect was not sustained at the three or six-month follow-up. This can indicate that consumers can revert to their prior purchasing trends after the initial "shock" of a 10% increase in the price of a sugary beverage.

A WHO report on taxing sugar-sweetened beverages (SSBS) found that a 20 percent tax would lead to an equivalent 20 percent decrease in consumption.[14] Further research on the effectiveness of this hypothetical 20 percent tax was conducted by Gardiner[15] and Finkelstein,[16] suggesting that higher tax rates, such as 20%, may reduce SSB consumption. In comparison, lower rates may have little to no impact on sales.

Research also finds that higher tax rates call for the re-formulation of beverage products in the sugar industry. [14,15,17] The application and severity of SSB taxes can vary across countries.

Types of Taxes:

Excise sugar taxes are usually placed on produced items and are sometimes imposed based on the amount of sugar (in grams) of a particular drink.[18] Excise taxes are also an example of a sin tax, which attempts to limit the consumption of a good that negatively affects society.

Value-added taxes are added at every production stage, including distribution and sale.[19] Each part of the production will typically be passed on to the next at a higher price to cover these costs. For example, when Coca-Cola ships out its bottles to retailers, the price of its distribution will be higher. When retailers (usually store owners) sell their bottles to consumers, they will charge a higher price to offset the tax. This effectively raises the price paid by the consumer. Ad valorem taxes are directly proportional to the product's price, meaning they are fixed percentage rates on beverages.[20] A higher-cost beverage would have a higher tax as a result.

Tax design plays an important role in reductions in obesity rates as seen through ad valorem taxes (percentage of a product, ex., 60 percent of the price is taxed), leading to larger reductions in total consumption compared to fixed tax rates, explained in The American Journal of Clinical Nutrition.[21] However, utilizing simulation methodologies, the research found that excise taxes were equally effective at reducing healthcare costs and preventing diseases as ad valorem taxes.[22]

Research advocates for specific taxes based on the amount of sugar rather than ad valorem taxes. Ad valorem taxes, which are calculated as a percentage of the product's price, can result in price equalization between sugary and diet beverages. This means that it reduces the motivation for consumers to choose healthier alternatives. Multiple researchers[8, 23-25] believe that the effectiveness of these taxes can be diminished if price equalization occurs.

Escobar[8] also supports the idea of taxing all sugar when it leaves the factory or enters the country. He argues that this approach might be more effective than targeting specific beverages, as it simplifies the implementation process and ensures everyone pays tax. Veitch[26] and Andreyeva[27] suggest that this unexpected trend may be attributed to price equalization strat-

egies, where retailers adjust the prices of diet and sugary drinks to minimize consumer switching based on price. This is mainly found in ad valorem taxes. It is important to note that the evidence on this specific topic is limited.

Substitution Effects of a Sugar Tax:

Research[4] found that although sugar-sweetened beverage consumption decreased, clear substitution effects may halt potential findings. Additionally, as multiple review papers[28-30] show, it is necessary to determine how much consumers may substitute their sugar consumption, calling for more extensive research on substitution effectiveness.

Studies done by Powell and Fletcher[31, 32] show that once caloric replacement and substitution are accounted for, there may be no net benefit of a tax on SSBS.

Research on Hungary, Mexico, France, and Polynesian sugar taxes found consumers switching to alternatives.[33] However, some researchers specify that the switch wasn't immediate, as seen in the Hungarian sugar tax.[34] Alternatively, some researchers have argued that consumers may not switch to alternatives. A 2020 study of substitution effects in Hungary found some, but confidence in predictions was very low.[35] The uncertainty stems from the observational study design, the potential misclassification of food products, and the specific parameters of the Hungarian tax.

Briggs also found that many people use other high-calorie or unhealthy products to substitute for their unhealthy eating when sugar-sweetened beverages are pricier.[36] Fowler also found substitution effects that may occur with various foods,[37] as a recent report has found pizza as a common substitute.

[38] This is significant as research has commonly found diet drinks as a substitute.[39] Aguilar[40] found that while states in the US have not experienced notable changes in substitution rates,[40] Mexico exhibited distinct substitution patterns. Within one year, the decrease in calorie and sugar consumption was counteracted by rises in the consumption of untaxed food items.

Ultimately, a recent meta-analysis[36] has hypothesized that a 10% tax on sugar taxes may slightly increase the consumption of untaxed beverages, such as water or milk. Still, this effect is not always significant and varies by jurisdiction.

Responses to Sugar Tax from Retailers:

Ponce's study of the Berkeley sugar tax found that in response to sugar taxes, many retailers raised prices of SSBS relative to non-taxed beverages.[40] Their study found that 84% of retailers raised the price of soda, and less than 5% increased the price of substitutes, such as water or milk.

Regressivity of a Sugar Tax:

Regressivity refers to low-income groups, which usually lack access to healthy foods due to their lower budgets and may struggle to find other adequate food/drink sources when these taxes are administered.[41] This can have a more pronounced impact on them than on a middle-or upper-income group.

Regarding analysis, Roberts points out the scarcity of peer-reviewed experimental and observational evidence restricting our understanding of the magnitude and the character of regressive (and progressive effects) sugar taxes.[42]

World Cancer Research Fund International recognizes the regressivity of such a tax and how it disproportionately affects lower-income populations.[43] However, the report argues that sugar-sweetened beverage (SSB) taxes are crucial to acknowledge due to the heavy burden of obesity and diet-related NCDS (Non-Communicable Diseases) within these communities.

Several studies have found regressive impacts of sugar taxes.[44-50] Governments need to assess whether a nationwide implementation of the sugar-sweetened beverage (SSB) tax benefits all socioeconomic strata, giving evidence supporting its primary impact on children, adolescents, and low-socioeconomic groups.

Some concerns have even been brought to policymakers, as certain nations have chosen to repeal sugar taxes over concerns regarding potential job losses. As the Department of Health, Ireland[51] study shows, these taxes can worsen inequality.

Research in South Africa[52] relates to Adam Smith's four principles for an effective tax, emphasizing the necessity for fairness. The equity of sugar taxes, or fairness, is debatable as it unfairly impacts low-income groups, who allocate a larger proportion of their income to sugary drinks.

Gruber and Kosegi[53] believe the benefits of corrective measures of sin taxes are indeed more pronounced for low-income groups. The correctness of a tax implies the ability to change behavior in both the short and long term, leading to healthier choices. This means that poorer consumers, who often exhibit a strong demand response to taxation, derive relatively greater advantages from taxes, even if the financial costs are regressive. Despite concerns about the prevalence of heart disease and other ailments, this does not imply that taxes are entirely detrimental to the well-being of low-income individuals, as the Journal of Economic Perspectives[10] agrees, along with alternative research.

However, some studies have not found significant regressivity issues. Research in Brazil[8] analyzing the Brazilian sugar tax found that, notably, SSBs constituted 1.7% of the total calories purchased by households, with their price roughly double that of other food and beverage items. Interestingly, while the relative consumption of SSBs tended to increase with income, peaking at 2.6% among the highest income quartile, the price per calorie of SSBs did not exhibit significant changes with rising income levels. The report found an income elasticity coefficient of 0.41, suggesting that a 1.00% increase in income would lead to a 0.41% increase in SSB consumption, implying the opposite effect where higher households may face the highest burden.

This evidence is also supported by a study in the UK,[36] which found reductions were more significant among higher-income groups than lower-income groups, contradicting the regressive trend in US modeling studies.

Similar research also found minimal differences,[55] finding that low-income individuals would pay slightly more SSB tax than their high-income counterparts, with per capita estimates at $19.97 and $18.84, respectively.

DOI: 10.36838/v7i10.155

Ultimately, conflicting reports on the regressivity of a sugar tax explain why regressivity can vary from location to location. However, it is important to note that much of the data finding non-regressivity was from the early 2010s.

It's important to note that many ad valorem taxes, such as the one in the UK, may only specifically apply to certain sugar-sweetened beverages. For example, they might only apply to "luxury" beverages or beverages already priced too high before a tax that low-income groups were not buying pre-implementation. As a result, high-income consumers may see more hard-hitting effects in those situations.

Conclusion

Contrary to popular belief, many studies have found that sugar taxes have inelastic demand, meaning consumers don't react swiftly to changing prices. This makes sense as the addiction to sugary beverages, regardless of the price, has contributed to the obesity pandemic.

Although specific price elasticity estimates may differ across studies, sugar taxes are seen to have primarily inelastic demand. This means that larger taxes will be more successful in curbing sugar consumption. Because of inelastic demand, consumers will bear more burden than suppliers when sugar taxes are implemented.

Overall, evidence suggests that well-designed sugar taxes—set at high rates—can reduce consumption, but may impact low-income individuals.

Acknowledgments

I truly thank Professor Aliyev for his tremendous support throughout the data-gathering process and for helping me identify the most relevant topics to focus on.

References

1. World Health Organization. Global action plan for the prevention and control of NCDs 2013-2020. 2013. Geneva: WHO. 2016.
2. American Heart Association. "How Much Sugar Is Too Much?" American Heart Association, 1 June 2023, www.heart.org/en/healthy-living/healthy-eating/eat-smart/sugar/how-much-sugar-is-too-much. Accessed 5 Sept. 2024.
3. Prater, M. "Price Elasticity: What It Is & How to Calculate It." Price Elasticity: What It Is & How to Calculate It, 25 Aug. 2022, https://blog.hubspot.com/sales/price-elasticity.
4. Colchero, M. A., Popkin, B. M., Rivera, J. A., & Ng, S. W. "Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study." BMJ, 2016c. https://doi.org/10.1136/bmj.h6704.
5. Paraje, G. (2016). The effect of price and Socio-Economic level on the consumption of Sugar-Sweetened Beverages (SSB): the case of Ecuador. PLoS ONE, 11(3), e0152260. https://doi.org/10.1371/journal.pone.0152260
6. Scarborough, Peter, Vyas, A., & Richard A. "Impact of the announcement and implementation of the UK Soft Drinks Industry Levy on sugar content, price, product size and number of available soft drinks in the UK, 2015-19: A controlled interrupted time series analysis." PLOS Medicine, 2019. https://pubmed.ncbi.nlm.nih.gov/32045418/.
7. Guerrero-López, C. M., Unar-Munguía, M., & Colchero, M. A. "Price elasticity of the demand for soft drinks, other sugar-sweetened beverages and energy dense food in Chile." BMC Public Health, 10 Feb. 2017, 17(1):180. https://doi.org/10.1186/s12889-017-4098-x.
8. 8. Escobar, M. A., Veerman, J. L., Tollman, S. M., Bertram, M. Y., & Hofman, K. J. "Evidence that a tax on sugar-sweetened beverages reduces the obesity rate: a meta-analysis." BMC Public Health, 2013, 13:1072. https://doi.org/10.1186/1471-2458-13-1072.
9. Claro, R. M., Levy, R. B., Popkin, B. M., & Monteiro, C. A. "Sugar-Sweetened beverage taxes in Brazil." American Journal of Public Health, 2012, 102(1), 178–183. https://doi.org/10.2105/ajph.2011.300313.
10. Pfinder, M., Heise, T. L., Boon, M. H., Pega, F., Fenton, C., Griebler, U., Gartlehner, G., Sommer, I., Katikireddi, S. V., & Lhachimi, S. K. "Taxation of unprocessed sugar or sugar-added foods for reducing their consumption and preventing obesity or other adverse health outcomes." Cochrane Library, 2020(4). https://doi.org/10.1002/14651858.cd012333.pub2.
11. Allcott, H., Lockwood, B. B., & Taubinsky, D. "Should we tax Sugar-Sweetened Beverages? An Overview of Theory and Evidence." The Journal of Economic Perspectives, 2019, 33(3), 202–227. https://doi.org/10.1257/jep.33.3.202.
12. 12. Cawley, J., Thow, A. M., Wen, K., & Frisvold, D. "The economics of taxes on sugar-sweetened beverages: a review of the effects on prices, sales, cross-border shopping, and consumption." Annual Review of Nutrition, 2019, 39, 317-338
13. Wansink, B., Hanks, A., & Just, D. (2012). From Coke to Coors: A field study of a fat tax and its unintended consequences. Science Research Network. https://doi.org/10.2139/ssrn.2079840
14. World Health Organization. *Health Taxes: A Primer*. Health Taxes: A Primer (a WHO policy brief). Switzerland: World Health Organization, 2019.
15. Gardiner, A. "Implications of a Sugar Tax in New Zealand: Incidence and Effectiveness." *Treasury*, 2016. http://www.treasury.govt.nz/publications/researchpolicy/wp/2016/16-09. Accessed 13 Oct. 2017.
16. Finkelstein, E. A., Chen, Z., Bilger, M., Nonnemaker, J., Farooqui, A. M., & Todd, J. E. "Implications of a Sugar-Sweetened Beverage (SSB) Tax When Substitutions to Nonbeverage Items Are Considered." *Journal of Health Economics*, 2013, 32(1), 219–239.
17. UNICEF, WHO, & World Bank Group. *Levels and trends in child malnutrition: UNICEF—WHO—World Bank Group joint child malnutrition estimates: key findings of the 2015 edition*. 2015.
18. Tax Policy Center. "How do state and local soda taxes work?" *Tax Policy Center*. https://www.taxpolicycenter.org/briefing-book/how-do-state-and-local-soda-taxes-work.
19. Cole, G. "Value-added tax: What is VAT and who has to pay it?" *Avalara, Inc.*, 7 Oct. 2023. https://www.avalara.com/blog/en/north-america/2021/07/what-is-vat.html.
20. "What is ad valorem tax? - Property Tax." *Tulare County*. https://tularecounty.ca.gov/propertytax/property-tax-accounting/faqs/what-is-ad-valorem-tax/.
21. Sassano, M., "National taxation on sugar-sweetened beverages and its association with overweight, obesity, and diabetes." *The American Journal of Clinical Nutrition*, 2019, 119(4), 990–1006.
22. Selimin, D. S., Safian, N., Hassan, M. R., Abdul Manaf, M. R., & Mohd Ghazali, Q. "Effectiveness of sugar sweetened beverages tax on health and healthcare costs: A systematic review." *International Journal of Public Health Research*, 2023, 13(2). https://spaj.ukm.my/ijphr/index.php/ijphr/article/view/460.23.
23. Bonnet, C., & Requillart, V. (2011). Does the EU sugar policy reform increase added sugar consumption? Empirical evidence on the

soft drink market. Health Economics, 20(9), 1012–1024. https://doi.org/10.1002/hec.1721
24. Brownell, K., Farley, T., Willett, W., Popkin, B., Chaloupka, F., Thompson, J., & Ludwig, D. "The public health and economic benefits of taxing sugar-sweetened beverages." *New England Journal of Medicine*, 2009, 361, 1599-1605
25. Wilson, N., Thomson, G., & Blakely, T. "Use of pricing and tax interventions for protecting health: potential relevance for New Zealand of recent international developments." *New Zealand Medical Journal*, 2011, 124(1342), 100-103.
26. Veitch J, Singh A, van Stralen MM, Van Mechelen W, Brug J, ChinAPaw Mai JM: Reduction in sugar-sweetened beverages is not associated with more water or diet drinks. Public Health Nutr. 2010, 14 (8): 1388-
27. Andreyeva, T, Cahloupka, FJ, and Brownell, KD. Estimating the potential of taxes on sugar-sweetened beverages to reduce consumption and generate revenue. Preventive Medicine 2011;52(6):413-416.
28. Public Health England. (2015). Sugar Reduction: the evidence for action. London.
29. Roberts, K.E., Ells, L.J., McGowan, V.J. A rapid review examining purchasing changes resulting from fiscal measures targeted at high sugar foods and sugar-sweetened drinks. Nutr & Diabetes 7, 302 (2017). https://doi.org/10.1038/s41387-017-0001-1
30. Le Bodo Y, De Wals P. Soda Taxes: The Importance of Analysing Policy Processes Comment on "The Untapped Power of Soda Taxes: Incentivising Consumers, Generating Revenue, and Altering Corporate Behaviours." Int J Health Policy Manag. 2018 May 1;7(5):470-473. doi: 10.15171/ijhpm.2017.126. PMID: 29764113; PMCID: PMC5953532.
31. Powell LL, Chriqui JF, Khan T, Wada R, Chaloupka FJ. Assessing the potential effectiveness of food and beverage taxes and subsidies for improving public health: a systematic review of prices, demand and body weight outcomes. Obesity Reviews 2013;14(2):110-128.
32. Fletcher JM, Fisvold DE, Tefft N. The effects of soft drink taxes on child and adolescent consumption and weight outcomes. Journal of Public Economics 2010;94(11-12):967-974.
33. Thow AM, Quested C, Juventin L, Kun R, Khan AN, Swinburn B. 2011. Taxing soft drinks in the Pacific: implementation lessons for improving health. Health Promot. Int. 26: 55–64.
34. Snowdon, Christopher, Sugar Taxes: A Briefing (January 13, 2016). Institute of Economic Affairs, Available at SSRN: https://ssrn.com/abstract=3852694
35. Pfinder M, Heise TL, Hilton Boon M, Pega F, Fenton C, Griebler U, Gartlehner G, Sommer I, Katikireddi SV, Lhachimi SK. Taxation of unprocessed sugar or sugar-added foods for reducing their consumption and preventing obesity or other adverse health outcomes. Cochrane Database Syst Rev. 2020 Apr 9;4(4):CD012333. Doi: 10.1002/14651858.CD012333.pub2. PMID: 32270494; PMCID: PMC7141932.
36. Briggs AD, Mytton OT, Kehlbacher A, Tiffin R, Rayner M, Scarborough P. Overall and income specific effect on prevalence of overweight and obesity of 20% sugar sweetened drink tax in UK: econometric and comparative risk assessment modelling study. BMJ. 2013 Oct 31;347:f6189. Doi: 10.1136/bmj.f6189. PMID: 24179043; PMCID: PMC3814405.
37. Finkelstein, Eric A., Chen Zhen, Marcel Bilger, James Nonnemaker, Assad M. Farooqui, and Jessica E. Todd. (2013). "Implications of a Sugar-Sweetened Beverage (SSB) Tax When Substitutions to Nonbeverage Items Are Considered." Journal of Health Economics 32(1): 219–39.
38. Aguilar A, Gutierrez E, Seira E. The effectiveness of sin food taxes: evidence from Mexico. J Health Econ. 2021;77:102455. doi:10.1016/j.jhealeco.2021.102455PubMedGoogle Scholar
39. Teng AM, Jones AC, Mizdrak A, Signal L, Genç M, Wilson N. Impact of sugar-sweetened beverage taxes on purchases and dietary intake: systematic review and meta-analysis. Obes Rev. 2019;20(9):1187-1204. doi:10.1111/obr.12868PubMedGoogle ScholarCrossref
40. Ponce J, Yuan H, Schillinger D, Mahmood H, Lee M, Falbe J, Daniels R, Madsen KA. Retailer perspectives on sugar-sweetened beverage taxes in the California Bay Area. Prev Med Rep. 2020 May 28;19:101129. Doi: 10.1016/j.pmedr.2020.101129. PMID: 32612904; PMCID: PMC7322349.
41. Kagan, J. (2024, February 6). Regressive Tax: Definition and types of taxes that are regressive. Investopedia. https://www.investopedia.com/terms/r/regressivetax.asp#:~:text=A%20regressive%20tax%20is%20one,percentage%20from%20high%2Dincome%20earners.
42. Roberts, K.E., Ells, L.J., McGowan, V.J. A rapid review examining purchasing changes resulting from fiscal measures targeted at high sugar foods and sugar-sweetened drinks. Nutr & Diabetes 7, 302 (2017). https://doi.org/10.1038/s41387-017-0001-1
43. World Cancer Research Fund International (2018). Building momentum: lessons on implementing a robust sugar sweetened beverage tax.
44. Tedstone A, Targett V, Allen R. Sugar reduction: the evidence for action. Sugar reduction: the evidence for action. 2015.
45. Schwendicke F, Stolpe M. Taxing sugar-sweetened beverages: impact on overweight and obesity in Germany. BMC Public Health. 2017;17(1):1-8.
46. Chudá T, Janský P. The impact of a fat tax: Progressive in health, but regressive in income? Prague Economic Papers. 2016;25(4):445-58.
47. Miao Z, Beghin JC, Jensen HH. Taxing sweets: sweetener input tax or final consumption tax? Contemporary Economic Policy. 2012;30(3):344-61.
48. Smith TA. Taxing caloric sweetened beverages: potential effects on beverage consumption, calorie intake, and obesity, DIANE Publishing; 2010.
49. Jou J, Niederdeppe J, Barry CL, Gollust SE. Strategic messaging to promote taxation of sugar-sweetened beverages: lessons from recent political campaigns. Am J Public Health. 2014;104(5):847–853.
50. Hartocollis A. Failure of state soda tax plan reflects power of anti-tax message. New York Times, July 3, 2010:Sect. A.
51. Department of Health, Ireland (2016). Introducing a tax on sugar sweetened drinks: health rationale, options and recommendations. [Online]. Available: http://health.gov.ie/wpcontent/uploads/2016/10/INTRODUCING-A-TAX-ON-SUGAR-SWEETENEDDRINKS-FINA.pdf. Accessed 2017/05/02.
52. Dams, J. (2017). EVALUATING THE EFFECTIVENESS OF SUGAR TAX TO COMBAT SOUTH AFRICAN'S SWEET TOOTH. International Journal of Economics and Finance Studies, 9(2), 83-98.
53. Gruber, J., & Kőszegi, B. (2004). Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes. Journal of Public Economics, 88(9–10), 1959–1987. https://doi.org/10.1016/j.jpubeco.2003.06.001
54. World Health Organization. Health Taxes: A Primer. Health Taxes: A Primer (a WHO policy brief). Switzerland: World Health Organization; 2019. p. 9.
55. Lin B., Smith, T. A., & Lee, J. (2010). The effects of a Sugar-Sweetened Beverage tax: consumption, calorie intake, obesity,

DOI: 10.36838/v7i10.155

and tax burden by income. RePEc: Research Papers in Economics. https://doi.org/10.22004/ag.econ.61167

Author

Vineet Vishnubhotla is currently a junior at Northwood High School and has a strong interest in applying data science and statistical models to economics to obtain practical insights.

IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

A Study on the Relationship between Adolescent SNS Overdependence and Adolescent Ego Identity

Yeongseo Lim

Global Vision Christian School BCA Broadfording, 13535 Broadfording Church Rd, Hagerstown, MD 21740, USA; yeongseolim59@gmail.com
Mentor: E Jeremijenko-Conley

ABSTRACT: Adolescents' SNS (Social Networking Service) usage greatly affects their relationships, mental health, and physical well-being. According to this study, adolescents tend to prefer their cyber identities (within SNS, video games, or other internet communities) to their "real" identity as a family member or student. This preference leads to smartphone immersion. This review article aims to examine the research surrounding adolescents' SNS usage, SNS overdependence, and the relationship between SNS overdependence and adolescents' ego identity. 2024 Statista demonstrates that the rate of smartphone ownership and usage among teenagers is higher than in other age groups, and is steadily increasing. The study also showed that female teenagers spend more time on social media than their male counterparts. The main purpose of SNS usage in teenagers was to communicate with friends or family (with TikTok and Instagram being the most popular mediums). Through analysis of preceding research papers, this literature review will demonstrate how adolescents' SNS overdependence is a central factor in ego identity formation. Developing self-control ability is the main factor in reducing SNS overdependence. This literature review will illuminate how to prevent adolescents' SNS overdependence and form an optimistic ego identity.

KEYWORDS: SNS (Social Networking Service), SNS Overdependence, Ego Identity, Adolescence, Smart Media.

Introduction

Recently, a new neologism, Phono sapiens, has emerged. This compound word of 'smartphone' and 'homo sapiens' refers to a new generation that conceptualizes smartphones as a part of their body, an extension of self.[1] Smartphones have become a widespread, deep-rooted element of adolescents' cultures.

According to the 2022 Pew Research Center,[2] nearly half of U.S. teens self-proclaim to use the internet "almost constantly." In 2014 to 2015, 95 percent of adolescents used the internet; however, in 2019, 97 percent of adolescents used the internet. The percentage (56%) of adolescents who use the internet several times a day and the percentage (46%) of adolescents who utilize the internet almost constantly both increased. This SNS overdependence has negative mental, physical, and relational effects.[4,9]

In adolescence, because of rapid changes in social, cognitive, and physical development, teenagers are often confused and threatened by a constantly shifting self-awareness. Therefore, the establishment of ego identity emerges as an important issue in adolescence.[10] With the development of Information and Communication Technologies, digital media techniques such as SNS expand opportunities to express one's identity beyond the familial or academic realm (whether that be through a virtual character or anonymous chat forum).

However, since SNS overdependence is a fairly recent development, research into the field is inadequate. Further, there is a dearth of research exploring the relationship between adolescents' SNS overdependence and their ego identity.

This review article aims to find characteristics of over-dependence through articles about SNS usage and other research. Moreover, it aims to analyze the relationship between adolescents' ego identity and SNS overdependence and provide the basic data for prevention programs that will encourage positive ego identity formation.

Key Words:

1. SNS (Social Networking Service)

SNS can be defined as a web-based online service where people reveal personal information publicly or privately to form relationships in the virtual space.[18]

2. SNS Overdependence

Smartphone overdependence refers to "a state in which excessive smartphone use increases the salience of smartphones and decreases self-control, resulting in serious consequences."[19,22]

Central to the definition of addiction is a dependence on a substance or activity. Smartphone addiction is generally conceptualized as a behavioral addiction. Literature suggests that mood tolerance, salience, withdrawal, and cravings have been associated with excessive smartphone use.[23]

In addition, in this study,[24] smartphone addiction means "a state of severe anxiety and loss of control due to excessive use and immersion in various functions of a smartphone, causing obstacles to his or her daily life." As described above, there is no significant difference in meaning between smartphone addiction and smartphone overdependence.

According to the Korea National Information Society Agency (NIA),[25] the tendency of SNS overdependence is similar to smartphone addiction or internet addiction. Also, SNS

DOI: 10.36838/v7i10.161

overdependence means continually using SNS even though one experiences serious psychological and social consequences because of increased salience and self-control failure.

Within the scope of this paper, SNS overdependence can be seen as similar to smartphone addiction or internet addiction. In addition, the paper will consider both the positive and negative aspects of internet usage, instead of treating adolescents as pathological diagnostic subjects. This study will utilize a unified word: SNS overdependence.

To sum up, this study defines SNS overdependence as using SNS constantly despite serious consequences, as well as experiencing withdrawal and tolerance by over-communicating and over-concentrating on SNS.

3. Ego Identity

Ego identity is the perception of selfhood: personality, future outcomes, and societal belonging. According to Erikson, the main developmental obstacle that adolescents face is ego identity confusion. In other words, the developmental task of an adolescent is to form a firm perception of who they are, where they are headed, and where they fit in.[26]

Sub-elements in ego identity can be divided into six categories: Independence, Self-Receptiveness, Future Certainty, Goal Orientation, Proactive Personality, and Intimacy.[27] First is independence, the degree to which one feels able to actively influence one's role and environment, also known as ability. Second is self-receptiveness, which is the self-perceived ability and talent level. Third is future certainty, the degree of hope one has for the future, the confidence to actualize one's dreams. Fourth is goal orientation, the will to fulfill tasks and realize conceptions of one's long-term identity.[27] Fifth is a proactive personality, the ability to recognize one's role and do tasks proactively.[29] The last category is intimacy, the level of group membership, degree of self-exposure, social flexibility (inhabiting various private and public selves), and the ability to maintain close relationships.

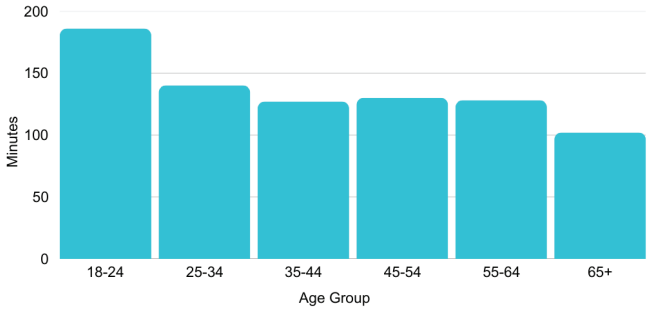
Methods

1. SNS usage status of Korean youth:

South Korea's rate of social media usage is second in the world. 92 percent of South Koreans use social media.[30] Especially among teenagers and young adults, smartphone retention rates were 100 percent and 99.4 percent, respectively, the highest among the age groups. The first and second most highly ranked purposes for smartphone usage are 'communication'.[31]

In the U.S., according to the February 2024 Statista **(Figure 1, Figure 2)**,[30] among several age groups: 18-24, 25-34, 35-44, 45-54, 55-64, and 65+, ages 18-24 showed the highest average time spent using social media. While users aged 18 to 24 spent 186 minutes a day on social media, respondents aged 65 and older utilized social media for 102 minutes daily. Preferred platforms also differ. TikTok was most used by teenagers aged 18-24. Their daily use averaged 76 minutes on the platform. Respondents aged 25-34, aged 35-44, and aged 45-54 spent most time on TikTok with an average of 50 minutes, 47 minutes, and 45 minutes, respectively. However, for respondents aged 55 to 65, Facebook dominated with approximately 45 minutes of daily usage.

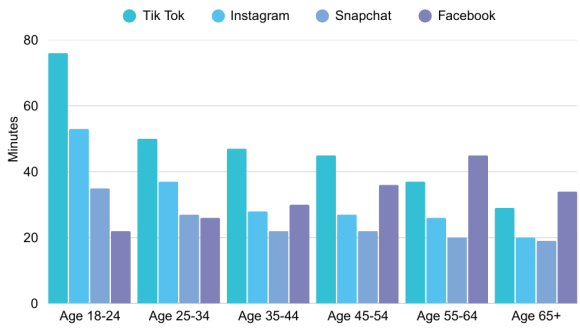
In South Korea, the 2022 Teenager Media Use Survey Report in Korea records that Korean teenagers' average daily internet usage is approximately 8 hours, more than their daily average of 7.2 hours of sleep. Today, the internet has consumed teenagers' daily lives, especially KakaoTalk and Instagram.



United States; February 2024; 18 years and older; Internet users who use a social network via any device at least once per month; includes all time spent on social network platforms; includes usage via any device

Total average time spent on social media daily by age group in the United States in February 2024

Figure 1: Average time spent on selected social media platforms daily among adults in the United States in February 2024, by age group (in minutes). Among all the age groupings, age 18-24 showed the highest average time spent using social media. (Created in Canva).



United States; February 2024; 18 years and older; Internet users who use a social network via any device at least once per month; includes all time spent on social network platforms; includes usage via any device

Average time spent on selected social media platforms daily among adults in the United States in February 2024

Figure 2: This table illustrates the favored social media platforms of various age groups. TikTok was most popular among teenagers and young adults aged 18-24. (Created in Canva).

According to the American Academy of Pediatrics (AAP),[32] teens frequently say that social media helps keep them connected to friends and family. This paper also found that teenage females tend to consider communication as more important than their male counterparts. Perhaps because of this, there are more female social media users than males. Therefore, this review article suggests that the government and other companies should devise programs to improve female students' relationships.

This study[55] showed that gender is a large factor in smartphone overdependence, game addiction, and SNS addiction.

The male students showed a higher tendency towards game addiction than their female counterparts. However, female students were more vulnerable to smartphone overdependence and SNS addiction than their male peers. In addition, this study showed that females with depression are more likely to become addicted to SNS than males struggling with depression.

2. The condition of adolescents' SNS overdependence:

This article[3] covers the condition of adolescents' SNS over-dependence through the Korea Ministry of Science and ICT and statistical data of the Korea National Information Society Agency (NIA), from September to November of 2023.

The Korea Ministry of Science and ICT and the Korea National Information Society Agency (NIA) collected data to prevent smartphone overdependence and find effective policy solutions. In 2023, researchers conducted 1000 South Korean household visits, surveying smartphone users aged 3-69. The collected data results are in **Table 1**. Table 1 shows the percentage of smartphone over-dependence risk groups in children aged 3 to 9, adolescents aged 10 to 19, adults aged 20 to 59, and older adults aged 60 to 69 from 2019 to 2023.

In 2019, the percentage of over-dependence risk groups was 22.9%. A 30.2% risk for smartphone over-dependence was shown in adolescents. In adulthood, the risk decreased to 18.8%, and shrunk further to 14.9% for those aged 60 to 69.

In 2023, the percentage of over-dependence risk groups in childhood was 25.0%. Adolescence was 40.1%, adulthood was 22.7%, and for those over 60, 13.5%. Adolescents' smartphone over-dependence was the highest among the age groups. In the last 5 years (2019-2023), the percentage of adolescents' smartphone over-dependence has increased by 32.7%.3 4 of 10 Korean adolescents are in danger of over-dependence.

Table 1: 2019-2023 Recent 5 Years Smartphone Over-dependence Risk Ratio.

Resource: The Report of 2023 The Condition of Smartphone Over-dependence from Korea National Information Society Agency (NIA).

Within the years 2019 to 2023, adolescents aged 10-19 showed the highest smartphone over-dependence among the age groups in each year 2019 to 2023. (Created in Canva).

Unit: %

Index	2019	2020	2021	2022	2023
Smartphone Overdependence Risk Group	20.0	23.3	24.2	23.6	23.1
Childhood (age 3-9)	22.9	27.3	28.4	26.7	25.0
Adolescence (age 10-19)	30.2	35.8	37.0	40.1	40.1
Adulthood (age 20-59)	18.8	22.2	23.3	22.8	22.7
Age 60 (age 60-69)	14.9	16.8	17.5	15.3	13.5

The three factors of smartphone over-dependence are self-control failure, resilience, and serious consequences. The result and scale of smartphone over-dependence are extracted from these three factors.[3]

First, self-control failure. Self-control failure means the users' control ability decreases due to smartphone usage compared to users' subjective goals.

Second is salience. Salience means that within one's life, using smartphones becomes one's most important and prominent activity.

The last is serious consequences. This refers to consistent smartphone usage despite the user's experience of negative physical, psychological, and social consequences.[3]

Each factor was rated on a scale of 4 points for prevalence in the adolescent over-dependence risk group. Self-control failure scored highest with 3.03, while salience was the second most prevalent (score 2.84), and serious consequences were lowest (score 2.44).

On the other hand, the general adolescent user group experienced self-control failure (score 2.10), salience (score 1.87), and serious consequences (score 1.72). Both groups showed that self-control failure was high for smartphone usage.

This NIA research concluded that excessive smartphone usage increases salience and decreases self-control ability, leading to serious consequences. This study[35] showed that social gratification and social media self-control failure (SMSCF) are two important mechanisms through which SNS use affects emotional health. Users may also experience negative emotions due to their internal failure to utilize self-control in regards to social media. The study illustrated that insufficient self-control in SNS use is closely associated with Facebook addiction and smartphone addiction. Another study found that users with low self-control tend to be attracted to friend-networking SNS, but not SNS with an entertainment and information sharing focus.[33] Other studies have shown that using SNS late at night is positively associated with sleep depletion, indicating a failure of self-control in using SNS during rest time.[33]

This study showed that self-control ability is the most impactful factor in adolescents' SNS overdependence.

3. The relationship between SNS overdependence and ego identity:

Adolescents are in the process of forming an ego identity, in other words, a concept of who they are and their place in society. This individuality is often found through a crisis of role confusion.[36]

Erikson[37] viewed ego identity as a continually improving sense of oneself, rather than a calcified social reality. Ego identity is not changed by external status, but through a gradual development of confidence in one's abilities to eventually arrive at internal homogeneity and continuity.

Waterman[38] defined ego identity as a clear concept of oneself, decisions about life goals, values, and beliefs, practical actions to carry out those decisions, degree of self-acceptance, a sense of one's own uniqueness, the exploration of several potential identities, and confidence in one's future. Weinreich[39]

DOI: 10.36838/v7i10.161

defined it as a self-understanding reached through reflection about one's past, present, and future.

The concept of ego identity by Erikson[36] suggests that it's the answer to the question 'Who am I?'. This constant self-evaluation is an essential component of maintaining a healthy personality.[27] If this ego identity cannot form, life development will be delayed, and negative symptoms such as low self-esteem and addiction will develop.[40] Studies have demonstrated a negative relationship between SNS addiction (similar to internet addiction or smartphone addiction) and ego iden-tity.[41-43]

- ***SNS activity has a positive effect on ego identity:***

Early adolescents require psychological time and space to face the task of identity formation. The smartphone space offers numerous possibilities to try out various selves.

Through SNS, adolescents can express themselves and communicate actively, utilizing texts, pictures, emoticons, and avatar nicknames. This paper showed that teenagers' immersion in smart devices has a positive effect on online communication. Within the virtual space, adolescents are able to engage in debates and emotional communication that they would not be comfortable with in "reality." This study[52] revealed a positive effect of online communication on one's ego identity.

This study[53] also showed that social media platforms can allow the sharing of one's interests and activities, through text, videos, and images. This type of personal sharing on social media, especially as an authentic form of self-expression, can help teenagers better understand who they are. It concludes that social media can help to build Identity, Self-Expression, and Self-Affirmation.

- ***SNS overdependence has a negative effect on ego identity:***

Middle school students plagued by internet addiction do not fully form an ego identity.[44] The communication addiction group possesses a low ego identity compared to the normal group.[42]

This study[45] shows that South Korean students' internet addiction has a significant impact on ego identity. There is a negative relationship between internet addiction and subfactors of ego identity, such as Self Receptiveness, Future Certainty, Goal Orientation, Proactive Personality, and Intimacy. According to preceding research from the U.S.,[17] the past experiences and behavior of users, and their ego-identities are important prediction factors in SNS usage. Also, ego identity and belongingness are highly impactful on the tendency towards SNS addiction.

This study[46] shows the degree of ego identity followed by excessive immersion in smart media. This study found that adolescents with excessive immersion in smart media show low ego identity.

This study[47] aims to examine the relationship between self-identity confusion and Internet addiction (IA) and the mediating effects of psychological inflexibility and experiential avoidance (PI/EA) indicators in college students. The researcher included that the severity of self-identity confusion was positively associated with the severity of PI/EA and IA. The results demonstrated that the severity of self-identity confusion was related to the severity of IA, either directly or indirectly. The indirect relationship was mediated by the severity of PI/EA. This research suggests that early detection and intervention of self-identity confusion and PI/EA should be the objectives for programs aiming to lower the risk of IA.

- ***The mediating effect of ego identity in the relationship between depression and aggression due to SNS overdependence:***

Adolescents' smartphone dependence has a high positive relationship with depression and other behavioral issues such as aggression. This study[48-50] also demonstrated a negative relationship between depression and aggression.

In this study,[51] adolescents' smartphone dependence affects depression, aggression, and ego identity was a mediating effect. The degree of depression and aggression in Korean adolescents who have high ego identity has been analyzed to be low depression and aggression so which shows a positive factor for adolescents.

Summarizing preceding research, **Figure 3** shows the relationship between SNS overdependence and adolescents' ego identity.

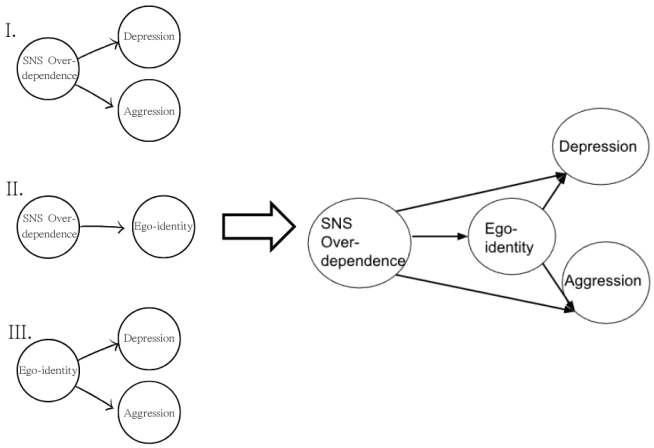


Figure 3: Summarizing preceding research, Figure 3 shows 'The Relationship Between SNS Overdependence and Adolescents' Ego Identity'. SNS Overdependence has a high positive relationship with depression and aggression. In turn, this depression and aggression can have a mediating effect on ego identity. (Created in Canva).

■ Conclusion

Through preceding research on the media usage of teenagers, studies conclude that the purpose of using SNS is to communicate with loved ones. In other words, teenagers highly prioritize communication in the digital world.

Also, SNS overdependence status reports show that teenagers' SNS overdependence is increasing, especially among teenage girls. Today, governments and other organizations are seeking several solutions to the issue of SNS overdependence. Governmental approaches to mitigating the digital addiction of minors have drastically different. For example, Australia's Kennedys law,[52] new legislation passed in 2024, prohibits children under 16 from using social media without verified parental consent. In the UK, 'The Online Safety Act,' set to

be fully implemented in 2025, enhances the protective framework by imposing a duty of care on online platforms to shield children from harmful content and interactions. Within the U.S., in 2024, California passed legislation to impose robust requirements on businesses offering online services likely to be accessed by children under 18.

But as these papers have shown, adolescents utilize SNS not only as a tool for communication, but also as an important way of performing and developing an identity. Put altogether, applying the benefits of using SNS, such as self-expression on SNS, forming a healthy culture among adolescents, and belongingness from communicating well, SNS overdependence prevention programs should be developed. This study[53] analyzed 13 research papers about the intervention methods of smartphone addiction. The results show that art therapy was found in 6 research papers (37.5%), exercise therapy in four research papers (25.0%), and cognitive behavior therapy in three research papers (18.6%). Future SNS overdependence prevention programs should include art therapy, as it has been demonstrated to mitigate SNS overdependence in adolescents. To maximize the function of communication, not only the government, but also companies should develop some programs or campaigns that can prevent SNS overdependence and reinforce belongingness and communication among friends.

The hypothetical research model proposed in this study—formulated to explain the causal relationships among adolescents' SNS overdependence, ego identity, aggression, and depression based on prior research and theoretical considerations—requires evaluation of its statistical adequacy through future empirical data collection.

Given that all variables within the model represent latent constructs and considering the importance of measurement validity and reliability, it is essential to employ statistical analysis methods based on structural equation modeling (SEM), which has been increasingly adopted in contemporary social science research.

■ Acknowledgments

I want to express my great gratitude to my mentor, Ms. E, for her help and support in completing this review article.

■ References

1. Naver.com. https://ko.dict.naver.com/#/search?range=all&query=%ED%8F%AC%EB%85%B8%20%EC%82%AC%ED%94%BC%EC%97%94%EC%8A%A4.
2. Vogels, E. A.; Gelles-Watnick, R.; Massarat, N. Teens, Social Media and Technology 2022. Pew Research Center. https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/.
3. [NIA National Information Society Agency]2023 Smartphone Over-dependence Survey Report. Nia.or.kr. https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=65914&bcIdx=26508&parentSeq=26508 (accessed 2025-01-01).
4. NIA. (2011). Development of a smartphone addiction diagnostic scale. Daegu: NIA
5. Kim. H.-S; Hwang J.-H. The Effects of Smartphone Dependence in Adolescents on Academic Helplessness: The Double Mediating Effects of Depression and Self-Esteem, Chonnam National University Graduate School, 2022. https://www.riss.kr/link?id=T16495181
6. Park, Y.-S. Influence of Adolescents' Smartphone Dependency on the Aggression Mediated by Depression. The Correction Welfare Society of Korea 2018, 55, 23–48. https://doi.org/10.35422/cwsk.2018.55.2.
7. Shin, E. Current status of smartphone usage and addiction among adolescents: Focusing on the body and mental health. RISS. https://www.riss.kr/link?id=T13317064
8. Kim, B.-N. Effect of Smartphone Addiction on Youth's Social Development. The Journal of the Korea Contents Association 2013, 13 (4), 208–217. https://doi.org/10.5392/jkca.2013.13.04.208.
9. Jo, S. The Relationship between Smartphone Addiction and Self-Esteem, Impulsiveness, and Sociability among High School Students. Domestic Master's thesis Sogang University, 2012, https://www.riss.kr/link?id=T12897811
10. Hyeonok, S.; Choi, S.-Y; Park, A.-C. A Structure Model Analysis between a Family Environment, Perceived Social Support, Self-Efficacy, and Ego-Identity of Adolescents. Korean Journal of Youth Studies 2010, 17 (2), 109–129. https://www.riss.kr/link?id=A104557920
11. Ellison, N. B., Heino, R., & Gibbs, J. (2006). Managing Impressions Online:- 97 Self-Presentation Processes in the Online Dating Environment. Journal of Computer-Mediated Communication, 11(2), 415-441.
12. Seong-ho, L.(2020). The Effects of Adolescents' Smartphone Dependency on Depression and Aggression: Focusing on the Mediating Effect of Self-Identity. Seoul Cyber University. https://www.riss.kr/link?id=T15624099
13. Oh, Y.-K. (2012). The Relationship Between SNS Addiction Tendency and Loneliness, Depression, Interpersonal Relationships, and Social Support. Master's Thesis, Korea Counseling Graduate School.
14. Koc, M., & Gulyagci, S. (2013). Facebook Addiction Among Turkish College Students: The Role of Psychological Health, Demographic, and Usage Characteristics. Cyber psychology, Behavior, and Social Networking, 16(4), 279-284.
15. Spraggins, A. (2009). Problematic use of online social networking sites for college students: Prevalence, Predictors, and association with well-being. Ph.D. University of Florida.
16. Wan, C. (2009). Gratifications & Loneliness as Predictors of Campus-SNS Websites Addiction & Usage Pattern among Chinese College Students. The Chinese University of Hong Kong.
17. Pelling, E. L. & White, K. M. (2009). The theory of planned behavior applied to young people's use of social networking websites. *Cyberpsychology, Behavior, and Social Networking*, *12*(6), 755-759
18. Choi S.-J; Lee J.-H. Constructing Media Rhythm Indices: Application of Post-TV Media Platforms. Journal of Communication Research. 2022, 59 (1), 139–184. https://doi.org/10.22174/jcr.2022.59.1.139.
19. 2019 Survey on Smartphone Overdependence, Available online: https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=65914&bcIdx=21939&parent Seq=21939
20. 2020 Survey on Smartphone Overdependence, Available online: https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=65914&bcIdx=23109&parent Seq=23109
21. 2021 Survey on Smartphone Overdependence, Available online: https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=65914&bcIdx=24288&parent Seq=24288
22. 2022 Survey on Smartphone Overdependence, Available online: https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=65914&bcIdx=25356&parent Seq=25356

DOI: 10.36838/v7i10.161

23. Ratan, Z. A.; Parrish, A.-M.; Zaman, S. B.; Alotaibi, M. S.; Hosseinzadeh, H. Smartphone Addiction and Associated Health Outcomes in Adult Populations: A Systematic Review. International Journal of Environmental Research and Public Health 2021, 18 (22), 12257. https://doi.org/10.3390/ijerph182212257.
24. S. B. Park and H. S. Hwang, "An Exploratory Study on Factors Influencing Smartphone Addiction: Focused on the Application Use of University Students", Korean Journal of Journalism & Communication Studies, vol. 58, no. 4, (2014), pp. 289-311.
25. Korea National Information Society Agency (2011). Development of a Smartphone Addiction Diagnosis Scale
26. Wallin, D. J. (2007). Attachment in Psychotherapy. Guilford Press.
27. Park, A.-C. Validation of Korean Adolescent Ego-Identity Scale. Korean Journal of Educational Psychology 2003, 17 (3), 373–392.
28. Kemp, S. (2023). Digital 2023: Global overview report. We Are Social. Young, K. S. (1999). Internet addiction: Evaluation and treatment. Bmj, 319(Suppl S4).
29. Korea Information Strategy Development Institute. Social Network Service(SNS) usage behavior by generation. KISDI STAT Report, 22-10. 2022
30. Statista. Time spent daily on social media U.S. by age 2024 | Statista. Statista. https://www.statista.com/statistics/1484565/time-spent-social-media-us-by-age/.
31. KoreaPressFoundation.www.kpf.or.kr. https://www.kpf.or.kr/front/board/boardContentsView.do?board_id=246&contents_id=29ff236264724e3fbe02e544185aac03.
32. Impact of Social Media on Youth. www.aap.org. https://www.aap.org/en/patient-care/media-and-children/center-of-excellence-on-social-media-and-youth-mental-health/qa-portal/qa-portal-library/qa-portal-library-questions/impact-of-social-media-on-youth/.
33. Liu, Y.; Yi, H.; Jiang, C. Enjoyment or Indulgence? Social Media Service Usage, Social Gratification, Self-Control Failure, and Emotional Health. International Journal of Environmental Research and Public Health 2023, 20 (2), 1002. https://doi.org/10.3390/ijerph20021002.
34. Erikson, E. H. (1968). *Identity: Youth and Crisis*. New York: Norton.
35. Erikson, E. H. (1982). *The life cycle completed*: A review. New York: Norton.
36. Waterman, A. S. (1982). Identity Development from Adolescence to Adulthood: An Extension of Theory and a Review of Research. *Developmental Psychology*, 18(3), 341- 358.
37. Weinreich, P. (1988). The operationalization of ethnic identity. Ethnic psychology. Research and practice with immigrants, refugees, native peoples, ethnic groups, and sojourners, 149-168.
38. Bynner, J., Ferri, E., & Shepard, A. (1997). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.
39. Ryu, Mi-ra (2002). The Relationship Between Adolescents' Internet Addiction, Social Support, and Self-Identity. Master's thesis, Graduate School of Education, Sogang University.
40. Seong, Young-Hye. Lee, Ju-Yeon. General Adolescents and Telecommunication Addiction Tendency Adolescents' Perception of Social Support and Self (2000). A comparative study on identity, SOOK-MYUNG JOURNAL OF CHILD STUDY, 14(2), 91-110.
41. Kim, Dae-eun. The Effects of Internet Addiction on the Establishment of Adolescent Self-Identity. Master's thesis, Graduate School of Education, Chongshin University. 2004
42. Park Ji-hye Lee. Dal-seok. A study on the Relationship between Internet Addiction and Identity of a Pupil in Middle school. , Vol 23 No.1 [2004]; Chosun University Student Guidance Research Institute, 2004
43. Kim, Nam-Soo (2011). The Effects of Internet Addiction on Adolescents' Self-Identity. Master's thesis, Graduate School of Public Administration, Kyunghee University.
44. Morita Yuka (2017). The Relationship Between Adolescents' Smart Media Use and Self-Identity. Master's Thesis, Graduate School, Sungkyunkwan University
45. Hsieh, K.-Y.; Hsiao, R. C.; Yang, Y.-H.; Lee, K.-H.; Yen, C.-F. Relationship between Self-Identity Confusion and Internet Addiction among College Students: The Mediating Effects of Psychological Inflexibility and Experiential Avoidance. International Journal of Environmental Research and Public Health 2019, 16 (17), 3225. https://doi.org/10.3390/ijerph16173225.
46. Koo Chang-mo (1992). Violence and Youth Problems. Seoul: Korea Youth Research Institute.
47. Shin, Min-Seop, and Oh, Kyung-Ja (1993). Identity confusion and maladaptive behavior in adolescence. Journal of College Life Studies, 1, pp.71-48.
48. Lim, Hye-sun (2016). A case study on self-expression and self-identity of maladjusted adolescents. Master's thesis, Graduate School of Education, Ewha Womans University
49. Lee, Sung-Ho. The Effects of Adolescents' Smartphone Dependency on Depression and Aggression: Focusing on the mediating effects of ego-identity; Department of Social Welfare, Graduate School of Human Service, Seoul Cyber University, 2020.
50. Gong, Moon-Jeong (2006). A study on the application of holistic education theory to the formation of positive identity of youth in cyberspace. Master's thesis, Graduate School of Education, Chosun University.
51. Shuna Shiann Khoo; Yang, H.; Toh Wei Xing. Leveraging Instagram to Enhance Self-Esteem: A Self-Affirmative Intervention Study and Multilevel Mediation Analysis. Computers in Human Behavior 2023, 150, 107972–107972. https://doi.org/10.1016/j.chb.2023.107972.
52. Moreno, D. N.; Pumphrey, B.; Anania, S.; Camilleri, M. 2025: A global shift towards digital regulation for children. Kennedys Law. https://kennedyslaw.com/en/thought-leadership/article/2024/2025-a-global-shift-towards-digital-regulation-for-children/ (accessed 2025-01-05).
53. Kim, D. A Systematic Review on the Intervention Program of Smartphone Addiction. 2020, 21 (3), 276–288. https://doi.org/10.5762/kais.2020.21.3.276.
54. Fleming-Milici, F.; Phrommavanh, P.; Phaneuf, L.; Harris, J. SUPPLEMENT: Examples of Social Media Campaigns Targeted to Teens and Hispanic and Black Youth; 2020. https://uconnruddcenter.org/wp-content/uploads/sites/2909/2020/09/SocialMediaCampaigns2020.pdf.
55. Kyoung, S.; Kim, J. A Study on Smart-Phone Addiction in Teenagers: Focused on Comparison Smart-Phone Overdependence, Game Addiction, SNS Addiction. Research Institute for Life and Culture Sogang University 2019, 52, 179–200. https://doi.org/10.17924/solc.2019.52.179.

Author

Yeongseo Lim is a junior at Global Vision Christian School BCA Broadfording in Maryland, the United States. Education and Psychology are my greatest academic passions. She aspires to work in educational engineering and hopes to develop an educational system.

IJHSR International Journal of High School Research

is a publication of

terra science and education

N.Y. based 501.c.3 non-profit organization
dedicated for improving K-16 education